THE KOREAN ECONOMIC REVIEW
Volume 19, Number 1, Summer 2003

AN EVOLUTIONARY ANALYSIS OF GAMES WITH PRE-PLAY COMMUNICATION*

YONG-GWAN KIM** · JOEL SOBEL***

"It takes two to tell the truth—one to speak and another to hear."
Henry David Thoreau

We consider a finite two-player game augmented by a round of pre-play communication. We show that if outcomes satisfy a stability condition suggested by adaptive dynamics, then pre-play communication effectively eliminates inefficient equilibria. We characterize the set of outcomes that satisfy stability conditions modeled after those used in biological game theory. A stable set of strategies is a closed set of Nash equilibria with the property that no other strategy can invade the population. If players have the same preferences over equilibria, then only the efficient equilibrium payoffs are stable when there is pre-play communication. We introduce a stronger notion of communication stability designed to capture the idea that introducing new words to the language should not destroy the stability of outcomes. A communication stable payoff must be an efficient point in the convex hull of the set of Nash equilibria, and any efficient element in the convex hull of Nash equilibria satisfying a regularity condition can be approximated by the payoff of some communication stable set.

JEL Classification: C62, C72, D82
Keywords: Evolutionary Stability, Pre-play Communication, Common Interest, Communication Stability, Efficiency Theorem.

I. INTRODUCTION

Although informal stories in game theory emphasize that pre-play communication allows players to coordinate on efficient Nash equilibria, these stories

Received for publication: Aug. 23, 2002. *Revision accepted:* Nov. 17, 2002.
* Kim thanks the Brain Korea 21 program financed by the Ministry of Education of Korea and Sobel is grateful to the Sloan Foundation and NSF for finacial support. Thanks also go to Antonio Cabrales, Vincent Crawford, Eddie Dekel-Tabak, Drew Fudenberg, Avi Shmida, Leo Simon, Jeroen swinkels, Joel Watson, seminar audiences at various institiutions and two referees for helpful comments. This paper is a revision of Kim and Sobel, "An Evolutionary Approach to Pre-play communication," 1992.
** Department of Economics, Sungkyunkwan University, Seoul, Korea.
*** Department of Economics, University of California, San Diego, USA.

are difficult to capture in full models of the communication process. The basic reason for the difficulty is that costless communication can never destroy a Nash equilibrium. If all but one player decides to ignore everything that is said and plays according to an equilibrium strategy, then the other player can do no better than speak randomly and also follow the equilibrium. Several authors have approached this problem by assuming that language exists and has a focal meaning obtained from its use outside the model. They continue by making behavioral assumptions that require players to believe the literal meaning of messages provided that these meanings do not violate strategic aspects of the game. Papers of Farrell [1988, 1993], Myerson [1983, 1989], and Rabin [1990, 1994] are examples of this work. Our approach is different. We do not assume that words have meaning outside the model. Instead, we show that if outcomes satisfy a stability condition suggested by adaptive dynamics, then pre-play communication effectively eliminates inefficient equilibria.

We add a round of pre-play communication to a finite two-player game. Each player simultaneously makes a statement from a finite language. The statements are revealed and then the underlying game is played. Pre-play communication is cheap talk in the sense that it does not directly enter the payoffs.

We characterize sets of strategies that satisfy stability conditions modeled after those used in biological game theory. A stable set of strategies will be a closed set of Nash equilibria with the property that no other strategy can invade the population. We use an entry condition due to Swinkels [1992] that describes the type of strategies that may invade. It requires potential invaders to use strategies that are optimal responses to the population strategy both before and after entry.

When we allow pre-play communication, the stability conditions work in the following way to move the population away from inefficient equilibria. Suppose the population plays a strategy that achieves a payoff which is dominated by another equilibrium payoff. If the possibilities for communication are sufficiently great, then there will be some redundant communication strategies: messages that are sent with probability zero in some element of the stable set. Invaders can use unsent signals to announce that they are prepared to play an efficient equilibrium in the underlying game, and to identify others similarly inclined. If the invader plays against a member of the original population, then its message will be ignored. The original population may interpret the new message in the same way as one of the words players are expected to use, so that there is no disadvantage to saying it. There is no reason for this invading strategy to die out. If the invader plays against another invader, then they coordinate on an efficient equilibrium, so there is an advantage to using the new strategy. The original population strategy is not stable.

In the next section we describe the communication game and motivate the solution concept. The third section describes a simple example. The fourth section discusses a preliminary result that plays a crucial role in the analysis. Section V presents an efficiency result for a class of common-interest games. If players have the same preferences over equilibria, then only the efficient equilibrium payoffs are stable when there is pre-play communication. Section V also contains an existence theorem for a related class of games. The sixth section explains why the efficiency result cannot be extended, even to the class of games in which there exists a unique efficient feasible payoff. Section VII introduces a stronger notion of communication stability, designed to capture the idea that introducing new words to the language should not destroy the stability of outcomes. In that section we present our main results: A communication stable payoff must be an efficient point in the convex hull of the set of Nash equilibria, and any efficient element in the convex hull of Nash equilibria satisfying a regularity condition can be approximated by the payoff of some communication stable set. We obtain these results without assuming that players have similar interests. Sections VIII and IX discuss the implications of assuming that players do not talk simultaneously. Section X discusses some related papers that use evolutionary arguments to select efficient outcomes in games.

II. THE FRAMEWORK

We begin with a given finite two-player game[1], which we call the underlying game. We add to the game one round of communication. Each agent has access to a finite set of messages (words) from a set M. For most of the paper we assume that players have access to the same set of messages and speak simultaneously. We discuss cases in which the players have asymmetric access to communication in Section VIII. The strategies of the players in the communication game are rules that specify a statement from M, and a function that maps the opponent's statement into the set of strategies in the underlying game. Payoffs for this game are precisely the payoffs obtained from the underlying game.

Formally, let the underlying game be (T, u), where $T = T_1 \times T_2$ (T_i is the finite strategy set for player i), and $u = (u_1, u_2)$. If player i uses stategy $t_i \in T_i$ for $i = 1$ and 2, then player j's payoff is $u_j(t_1, t_2)$. In the communication game with message space M, player i's strategy space is $M \times T_i^{M \times M}$, and the payoff to player j if player i uses strategy $(m_i, t_i(\cdot)) \in M \times T_i^{M \times M}$ is

$$U_j(m_1, t_1(\cdot), m_2, t_2(\cdot)) = u_j(t_1(m_1, m_2), t_2(m_1, m_2))\ ^2. \tag{1}$$

---

[1] We limit attention to two-player games because pairwise contests have been the standard setting for evolutionary games. If we assume that all messages can be heard by all players, then our results extend in a natural way to games with many players.

[2] In the reduced normal form of the communication game a player need not plan a response to messages that it does not send. We can represent the pure strategies available to player i by $M \times T_i^M$, where player i's strategy $(m_i, t_i(\cdot))$ specifies that the player will send the message

We denote the communication game with message space M and underlying game (T, u) by (M, T, U), where $U = (U_1, U_2)$ and (1) defines U_j. In the usual way, we extend payoff functions to mixed strategies using linearity.

Talk is cheap since the messages do not enter payoff functions directly. Messages influence payoffs only to the extent that they affect the actions players choose in the underlying game.

For this class of games there is always a Nash equilibrium in which players play the same actions in the underlying game for all communication histories. Nash equilibrium cannot force particular statements to have meaning unless the statements are directly linked to payoffs. Nevertheless, allowing pre-play communication changes the strategic environment in a fundamental way. Strict Nash equilibria in the underlying game are no longer strict equilibria in the communication game. This observation suggests that processes which use communication to move the population gradually away from inefficient equilibria could evolve. The outcomes that we study, which satisfy stability conditions suggested by evolutionary processes used in biological game theory, cannot ignore cheap talk.

We work with a static stability condition that substitutes for a full description of the evolutionary process. The original stability condition of this sort is the notion of evolutionarily stable strategy (ESS) of Maynard Smith and Price [1973]. This concept is poorly suited for application to extensive games (see Selten [1983] or Swinkels [1992] for discussions of the problems). We use a modification of ESS developed by Swinkels, who requires that strategies be stable only against invasions that respond optimally to the perturbed environment. We state the definition for a general two-player game with strategy set $S = S_1 \times S_2$ and payoff functions $\pi = (\pi_1, \pi_2)$, which we represent by (S, π). Let $N(S, \pi)$ be the set of Nash equilibria of (S, π); let $\Delta(S_i)$ be the set of mixed strategies of player i; let $C(\sigma)$ be the carrier of σ(the set of pure strategies given positive probability by σ); let $BR_i(\cdot)$ be the best response correspondence of player i for $i = 1$ and 2; and for $\sigma = (\sigma_1, \sigma_2) \in \Delta(S_1) \times \Delta(S_2)$, let $BR(\sigma) = (BR_1(\sigma_2), BR_2(\sigma_1))$.

Definition. A set $\Theta \subset \Delta(S_1) \times \Delta(S_2)$ is **equilibrium evolutionarily stable** (EES) if it is minimal with respect to the following conditions:

$$\Theta \text{ is closed and nonempty.} \tag{i}$$

$$\Theta \subset N(S, \pi). \tag{ii}$$

There exists $\varepsilon' \in (0, 1)$ such that for all $\varepsilon \in (0, \varepsilon')$ and for all $\sigma \in \Theta$, if $C(\sigma') \subset BR((1-\varepsilon)\sigma + \varepsilon\sigma')$, then $(1-\varepsilon)\sigma + \varepsilon\sigma' \in \Theta$. (iii)

---

m_i and respond to the message m_j with $t_i(m_j)$. To simplify our notation we include a description of redundant strategies. Our results do not depend on which of these specifications we use.

Conditions (i) and (ii) in the definition are familiar. They require that Θ be a closed set of Nash equilibria. Condition (iii) is novel. It is the invasion condition. It states that if σ is in the stable set, and σ' responds optimally to the perturbed environment, then the population average strategy following the invasion is also in the stable set. Swinkels [1992] provides several motivations for condition (iii).

The following informal story describes a process that justifies evolutionary stability conditions. Two role-identified (row players and column players) populations play in pairwise contests. The outcome of each contest determines a players payoff. Participants play a fixed (mixed) strategy. Established members of the population die regularly and are replaced either by another individual who plays the same strategy or an invader. Invaders first enter the population on a probationary basis. They survive if and only if they perform at least as well as the rest of population for a finite interval following their arrival. Hence they must not only do as well as the established population, but they must do as well as any new invader that may follow it. Such a condition seems appropriate if some members of the general population can change their strategy in response to a perturbation faster than the invaders themselves or if there exists a pool of potential entrants who are able to respond optimally to any population strategy. When condition (iii) holds, σ' survives its probationary period against a population strategy σ. An EES set is a minimal set with the property that no invasion from outside the set can survive.

There are three important differences between the ESS and EES stability conditions. First, the basic notion of ESS assumes that the underlying game is symmetric, and that roles players take (column or row) are not identified. We choose to concentrate on games in which roles are identified primarily because it allows us to discuss all two-player games without resorting to symmetrization.

Second, EES is a set-valued concept. ESS will never exist in games with nontrivial opportunities for pre-play communication. The example in Section III will make this clear.

Third, admissible invasions in the EES framework must respond optimally to the population mixture that they induce. Without this restriction, pre-play communication would disrupt the (Defect, Defect) equilibrium in the prisoner's dilemma since a strategy that used different messages could invade the population, play the Nash equilibrium against the existing population, and cooperate with each other.[3] Players with even limited ability to forecast future behavior would not adopt this invasion strategy because they would realize that a new group of invaders could take advantage of them by imitating their signal and

[3] Robson [1990] considers the possibility of creation of extra strategies in evolutionary games. These strategies play the same role communication does in our model. Robson demonstrates how adding a strategy forces cooperation in coordination games and destabilizes the inefficient outcome in the one-shot prisoner's dilemma. He argues that the creation of further strategies in the prisoner's dilemma will restore the stability of the (Defect, Defect) outcome.

then defecting. While we believe that condition (iii) is plausible in our framework, it is not needed to prove our existence and efficiency results. Rather it guarantees existence of stable sets in games like the prisoner's dilemma with pre-play communication.[4]

Swinkels [1992] shows that EES sets have several attractive properties. For example, every EES set contains a proper equilibrium and satisfies the never-weak-best-response property of Kohlberg and Mertens [1986]. EES sets continue to be EES sets after weakly dominated strategies are deleted from the game and (unlike ESS) are robust to the inclusion or exclusion of redundant strategies.

Even though EES is substantially weaker than ESS, there is no guarantee that EES sets exist. Swinkels presents an example, and in Section V we provide a non-pathological example of a game in which pre-play communication destroys all stable sets of the underlying game.

To obtain a general existence result condition (i) must be abandoned. Gilboa and Matsui [1991] do this with their concept of cyclically stable sets. Gilboa and Matsui show that cyclically stable sets, which satisfy a condition similar to (iii) but need not satisfy condition (ii), generally exist. Matsui [1991] applies the concept of cyclically stable sets to pre-play communication and obtains results similar to ours. We discuss this work in Section X.

We use the following proposition to characterize EES sets.

Proposition 1. Let Θ be an EES set. If $\sigma' \in N(S,\pi)$ and $C(\sigma') \subset BR(\sigma)$, then $(1-\lambda)\sigma + \lambda\sigma' \in \Theta$ for all $\lambda \in [0,1]$.

Proof. Let $\sigma' \in N(S,\pi)$ and $C(\sigma') \subset BR(\sigma)$. It follows that

$$C(\sigma') \subset BR((1-\lambda)\sigma + \lambda\sigma') \text{ for all } \lambda \in [0,1] \tag{2}$$

Let $\varepsilon'' = \sup \{\varepsilon' : (1-\varepsilon)\sigma + \varepsilon\sigma' \in \Theta \text{ for all } \varepsilon \in (0,\varepsilon')\}$. By (iii), $\varepsilon'' > 0$, and by (i), $\sigma'' \equiv (1-\varepsilon'')\sigma + \varepsilon''\sigma' \in \Theta$. From (2) it follows that $C(\sigma') \subset BR((1-\lambda)\sigma'' + \lambda\sigma')$ for all $\lambda \in [0,1]$. Therefore (iii) implies that $\varepsilon'' = 1$ and extablishes the proposition.

Ⅲ. EXAMPLE

Before proceeding with the formal analysis, in this section we discuss a

[4] Cressman [1992] and Thomas [1985a and b] analyze a set-valued, symmetric version of evolutionary stability that does not impose a restriction on entrants. This definition can be adopted to asymmetric games without difficulty to obtain a stability notion that is more difficult to satisfy than EES. Therefore our existence results would hold for this definition. It is straightforward to show that our efficiency results would also continue to hold.

simple example in order to illustrate the nature of our results. Consider the game in Figure 1. This coordination game, which Binmore [1992] calls DODO, has two pure-strategy Nash equilibria and a completely mixed one. There is no conflict of interest in this game and there are many reasons to expect that experienced players would coordinate on the efficient equilibrium. The inefficient pure-strategy equilibrium is a problem for standard theory, however. If for some reason a player believes that other players will be playing BAD, they receive their highest payoff only if they play BAD themselves.

[Figure 1]

	GOOD	BAD
GOOD	2,2	0,0
BAD	0,0	1,1

Without pre-play communication it is difficult to see how players can move away from the inefficient strict equilibrium since a unilateral deviation from the equilibrium strategy leads to a strict decrease in payoff. If pre-play communication is possible, moving to the efficient outcome is possible. When players have more than one communication strategy before they reach the underlying game, the communication game will have no strict equilibrium. To see this, focus on an equilibrium in which players receive the payoff one. Distinguish two cases depending on whether exactly one pair of messages receives positive probability in the equilibrium or more than one pair of messages receives positive probability in the equilibrium. In the first case, there will be many ways to respond to unsent messages; in the second case, there will be at least one player who sends more than one message with positive probability; such a player must be indifferent between which message it sends in equilibrium; it may vary the probabilities that it sends either message and still respond optimally. Games with pre-play communication lack strict equilibria, so they will typically fail to have ESSs; this is why we use a set-valued solution concept.

We now explain how our solution forces efficiency in this example. Failure to play a Nash equilibrium is not evolutionarily stable even without the possibility of pre-play communication. Possibly the population is playing the mixed-strategy equilibrium to the underlying game after some messages. In DODO this behavior is not stable as a group of invaders that always plays the GOOD strategy will be viable: When the population tilts slightly towards the efficient equilibrium, it becomes uniquely optimal for more invaders ready to play GOOD to enter. This argument does not depend on whether pre-play communication is possible; it is the requirement that outcomes be stable against entrants alone, and not pre-play communication, that destroys the mixed-strategy equilibrium in

DODO.[5]

Suppose that the population has coordinated on an equilibrium in which the players always use a particular (normal) message. Provided their opponent uses that message, they play the BAD action. Otherwise (off the equilibrium path) they do something that yields a payoff strictly less than one to anyone who uses an abnormal message (this can be done, for example, if the response to an abnormal message is play GOOD with probability one third and BAD with probability two thirds). Against this population strategy it is not profitable for an invader to use an abnormal message. The population punishes those who use unfamiliar words. As long as abnormal words are not used, however, there is no pressure to respond to them in a particular way. One generation of invaders that appears to play exactly like the general population can enter the population. These invaders always send the normal message and play the BAD strategy in the underlying game no matter what the other player says. While these invaders are no better than the general population, they are also no worse. Our solution concept implicitly assumes an environment where there are repeated possibilities for invasion, so all potential neutral invasions should occur in the long run. Hence the population can drift towards a configuration in which players who use an abnormal message are not punished.

Now consider an invading strategy that sends an abnormal message, and plays the GOOD strategy in the underlying game if and only if it meets another player who sends an abnormal message. This invader does not lose anything when it plays the original population. It strictly gains when it plays another invader. Consequently, it thrives.

We assumed in our discussion that there was a point at which everyone in the population sent the same message. Given this assumption a player could use an abnormal message to signal that it was willing to play the GOOD strategy in the underlying game. For coordination games such as DODO it is possible to prove that in any stable set there will be a strategy in which some message is not sent. If the population were playing a strategy that used all messages with positive possibility, then each player would be indifferent between the messages that it chooses to use against the existing population. Moreover, both row and column player can agree (in DODO) about which pair of messages lead to the best outcome in the underlying game. An invasion of players who use a specific message pair that leads to the favorite outcome can enter the population.

Pre-play communication in our framework only forces efficiency when we can guarantee that every EES set contains a strategy in which there is an unused message; guaranteeing this property requires two assumptions. First, there must be no conflict of interest between the players in the underlying game. If players do not have the same preferences over the equilibria in the underlying game,

[5] In Example 1 playing the mixed-strategy equilibrium is ruled out by our solution concept; in general mixed strategies may be elements of EES sets.

then an invader using a pure signaling strategy may provide its opponent an opportunity to take advantage of it. We discuss an example with this property in Section VI.

We must also assume that the game is played by two role-identified populations. Without this assumption we could not guarantee the existence of an unused message. Pre-play communication does not lead to efficiency under a single-population definition of stability even in a symmetric coordination game like DODO. Consider the strategy that randomizes equally between each of two messages, and then plays BAD if and only if its message is the same as the message of its opponent. When everyone in the population uses this strategy, the expected outcome of each meeting is coordination on the (GOOD, GOOD) equilibrium of DODO with probability one half, and coordination on the (BAD, BAD) equilibrium with probability one half. While a strategy that sends a message with probability one (and plays appropriately in the underlying game) is an optimal response to the population strategy, it performs poorly against another strategy that behaves in the same way because when matched together their messages agree with probability one, so they always coordinate on the (BAD, BAD) equilibrium. Consequently the random strategy is, taken as a singleton, an EES set (and even an ESS) for the communication game when viewed as being played by a single population. In the two-population version of DODO, an invading strategy profile that uses a message pair leading to the best equilibrium reached by the population strategy can always enter. Since the invader's message pair may be asymmetric, the same conclusion does not follow in the single-population version of the game.

Our argument demonstrates that inefficient outcomes are not evolutionarily stable when we add pre-play communication to DODO. The argument hinges on the way in which our solution concept permits strategies to change off the equilibrium path. Section IX discusses alternative models in which this type of drift arises, and the implications of evolutionary solution concepts that do not permit drift.

IV. THE BASIC LEMMA

In this section we present the basic lemma which demonstrates that if there exists an element of an EES set that does not use a particular signal, then there is an element of the EES set that also does not use the signal, but for which there is an optimal response that does use the signal. That is, a player has nothing to lose from using an unsent message.

In what follows we represent a (mixed-) strategy profile of the communication game by a pair (μ, τ), where $\mu=(\mu_1, \mu_2)$ and $\mu_i(m_i)$ is the probability that player i sends the message m_i, and $\tau=(\tau_1, \tau_2)$ and $\tau_i(t_i; m_1, m_2)$ is the probability that player i takes the action t_i in the underlying game following the message pair (m_1, m_2).

Lemma. Let Θ be an EES set, If there exists $\sigma=(\mu,\tau)\in\Theta$ and $(\overline{m}_1, \overline{m}_2)\in M\times M$ shch that $\mu_i(\overline{m}_i)=0$ for $i=1$ and 2, then there exists $\sigma'=(\mu',\tau')\in\Theta$ such that $U(\sigma)=U(\sigma')$ and $\mu_i'(\overline{m}_i)=0$ for $i=1$ and 2 and $t_i\in T_i$ such that $(\overline{m}_i, t_i)\in BR_i(\sigma_j')$ for $j\neq i$ and $i=1$ and 2.

The proof of the lemma is in the appendix. We provided the intuition for the result in our discussion of the example. If the population is playing a strategy that never uses a message, then there is nothing to prevent the population's response to the message from drifting until players interpreted the message exactly the same way as a word used with positive probability.

V. GAMES WITH (EQUILIBRIUM) COMMON INTEREST

In this section we study pre-play communication when the players have similar preferences in the underlying game. We study two types of games. In one class of underlying games the players have the same preferences over equilibria. When we add pre-play communication to these games of equilibrium common interest the only possible EES payoff is the player's most preferred equilibrium payoff. In the other class of common-interest games the feasible set of payoffs has a unique efficient point. For underlying games with this property we show that there exists an EES set for the communication game that attains the efficient payoff. We begin with a discussion of the two definitions. We then state and prove the efficiency theorem and the existence theorem. We conclude the section with an example of a communication game with no EES set.

A game (S,π) has **equilibrium common interest** (ECI) if s and $s'\in N(S,\pi)$, then for i and $j=1$ and 2, $\pi_i(s)>\pi_i(s')$ implies that $\pi_j(s)>\pi_j(s')$. That is, both players have the same rankings over Nash equilibria.

The game (S,π) has **common interest** (CI) if $\pi(S)$ has a unique weakly efficient point. That is, there exists $\pi^*=(\pi_1^*,\ \pi_2^*)$ such that

$$\pi(s^*)=\pi^* \text{ for some } s^*\in S, \tag{3}$$

$$\pi(s)\leq\pi^* \text{ for all } s\in S, \text{ and} \tag{4}$$

$$\pi_i(s^*)=\pi_i^* \text{ implies } \pi(s)=\pi^* \text{ for all } s\in S \text{ and } i=1 \text{ and } 2. \tag{5}$$

Condition (3) states that π^* is feasible; condition (4) states that it is efficient; and condition (5) states that it is the only efficient point. Plainly $s^*\in N(S,\pi)$.

Common-interest games are a natural place to look for effective communication. However, we show by example in the next section that the assumption of common interest is not sufficient to rule out inefficient EES sets. Instead we must use the assumption of equilibrium common interest to rule out inefficient outcomes. By focusing only on equilibria, there is a sense in which ECI is less restrictive than CI. For example, any game with a unique Nash equilibrium has equilibrium common interests. Moreover, if a game has equilibrium common interests, then there may exist feasible outcomes (which are not equilibrium outcomes) that both players prefer to their most preferred equilibrium outcome. The prisoner's dilemma is an example of a game that satisfies ECI but not CI. On the other hand, CI is a less restrictive assumption than ECI in the sense that players need not have the same preferences over inefficient equilibria. At the end of this section we discuss an example of a game with equilibrium common interest but without common interest. In the next section we give an example of a game that satisfies CI but not ECI.

Throughout this section we assume that (T, u) satisfies either ECI or CI and in either case we denote by $u^* = (u_1^*, u_2^*)$ the payoffs of the most preferred equilibrium.

First we show that stable sets in games with equilibrium common interests must yield the efficient equilibrium payoff.

Proposition 2. Assume that (T, u) has ECI and M has at least two elements. If Θ is an EES set of (M, T, U), then $U(\sigma) = u^*$ for all $\sigma \in \Theta$.

We present a proof of the proposition in the appendix. The proof is in two steps. First, we show that there exists an element of Θ that does not use all messages. To show this we start with an $\sigma \in \Theta$. If under σ all messages are sent with positive probability, then the strategy induces Nash equilibrium behavior in the underlying game after all messages. By the ECI assumption, players have the same preferences over these equilibria. Consequently, a strategy that uses only a message pair leading to the best of these equilibria can enter the population. Second, we apply the lemma to show that there exists a $\sigma' \in \Theta$ in which players can use an unused message without being punished. We use this strategy to construct an element of Θ that yields the efficient payoff, and to show that all elements of Θ yield this payoff.

Since there is no guarantee that EES sets exist in all situations, Proposition 2 is of limited interest without an associated existence result. Proposition 3 proves that EES sets exist in common-interest games.

Proposition 3. If (T, u) has common interest and M contains at least two elements, then $\Theta = \{\sigma : U(\sigma) = u^*\}$ is an EES set of (M, T, U).

Proof. Θ is a closed subset of Nash equilibria. If $C(\sigma') \subset BR(\sigma)$, then

$$U_1(\sigma_1', \sigma_2) = U_1(\sigma) = u_1^* \text{ and } U_2(\sigma_1, \sigma_2') = U_2(\sigma) = u_2^*. \quad (6)$$

It follows from (5) and (6) that

$$U(\sigma_1', \sigma_2) = U(\sigma) = u^* \text{ and } U(\sigma_1, \sigma_2') = U(\sigma) = u^*. \quad (7)$$

To show that Θ satisfies condition (iii) in the definition of EES, it suffices to show that $U(\sigma') = u^*$ whenever there exists $\varepsilon > 0$ such that $C(\sigma') \subset BR((1-\varepsilon)\sigma + \varepsilon\sigma')$.
However, it follows from the second pair of equations in (7) that

$$U_1(\sigma_1, (1-\varepsilon)\sigma_2 + \varepsilon\sigma_2') = u_1^*. \quad (8)$$

Hence, to satisfy the invasion condition, it must be that

$$U_1(\sigma_1', \ (1-\varepsilon)\sigma_2 + \varepsilon\sigma_2') = u_1^*. \quad (9)$$

(9) can only hold if $U_1(\sigma_1', \sigma_2') = U_1(\sigma) = u_1^*$.
Applying the same reasoning to the second player leads to the desired result.
We omit the details of the routine verification showing Θ is a minimal set that satisfies the conditions of the definition.

The stable set Θ described in Proposition 3 consists of all strategies with the property that at least one pair of signals leads to the efficient equilibrium, and that players always choose signals that lead to the efficient payoff. If M had only one element, then there would still exist an EES set with payoff u^*, but the set could be a proper subset of Θ. This is the case in common-interest games that have multiple efficient equilibria.

While in general CI does not imply ECI, we can derive the following corollary for 2×2 games by combining the proofs of Propositions 2 and 3.

Corollary. if (T, u) is a 2×2 game with common interest and M has at least two elements, then $\{\sigma : U(\sigma) = u^*\}$ is the only EES set of (M, T, U).

For general games, the ECI and CI assumptions combined are extremely restrictive. They hold for the class of pure-coordination games, in which the players have identical preferences. Nöldeke, Samuelson, and van Damme [1991] and Wärneryd [1991] study the effect of pre-play communication and evolution in this class of game.

The assumption of common interest in the existence theorem is a strong one. The next example suggests that it is difficult to weaken. The game in Figure 2 has two strict equilibria (UP, LEFT) and (DOWN, MIDDLE), and a mixed-strategy equilibrium ((.8, .2), (.75, 0, .25)) with payoffs (7.5, 8). The players have the same preferences over equilibria, so the game has ECI. It is not a game of common interest.

[Figure 2]

	LEFT	MIDDLE	RIGHT
UP	10, 10	0, 0	0, 8
DOWN	5, 0	10, 10	15, 8

We will show that no EES set exists when there is pre-play communication prior to the play of the game in Figure 2. Let M have two elements, α and β; enlarging the message space would not change the example. Describe a players pure strategy for (M, T, U) as a triple (ξ, X, Y) where ξ is the message, X is what it plays if its opponent says α, and Y is what it plays if its opponent says β. It is straightforward to check that if there is an EES set for (M, T, U), then

σ=((α, UP, DOWN), [.5(α, LEFT, LEFT), .5(β, MIDDLE, (.5,0,.5))])

must be in the set. Under σ half of the time both players say α and they coordinate on the (UP, LEFT) equilibrium. The other half of the time the message pair is (α, β) and the players coordinate on the (DOWN, MIDDLE). Now consider the strategy

σ'=((β, UP, DOWN), (β, MIDDLE, MIDDLE)).

This strategy is an efficient Nash equilibrium of the communication game. Furthermore, σ' is an optimal response to all mixtures of σ and σ'. Consequently, for sufficiently small positive ε, $(1-\varepsilon)\sigma+\varepsilon\sigma'$ must also be in any EES set. Since $(1-\varepsilon)\sigma+\varepsilon\sigma' \notin N(M, T, U)$ for all $\varepsilon \in (0,1)$, no EES set can exist.

In the game described in Figure 2 there are two strict Nash equilibria. Since the game satisfies ECI we know that any EES set must contain an efficient equilibrium. When we add pre-play communication the population can move from a state in which it coordinates on the (UP, LEFT) equilibrium (using σ) to a state in which it coordinates on the (DOWN, MIDDLE) (using σ') equilibrium. Hence any EES set for the game must contain both of these strategies. In addition, our argument shows that any EES set containing σ must contain strategies of the form $(1-\varepsilon)\sigma+\varepsilon\sigma'$; since these are not Nash equilibria (σ_2 dost not specify an optimal response to β), no EES set exists.

While the example suggests that it is difficult to have an existence theorem that applies to a wide class of games with pre-play communication, we show in Section Ⅶ that limitations on the set of available messages leads to existence of EES sets under relatively weak conditions.

Ⅵ. ARGUMENTS: HOW PRE-PLAY COMMUNICATION FAILS TO GUARANTEE EFFICIENCY

Proposition 2 applies to a limited class of games. One might hope that at least it extends to common-interest games. The next example demonstrates that a more general result cannot be obtained.

[Figure 3]

	BALLET	FIGHT	BED
BALLET	2, 1	0, 0	0, 0
FIGHT	0, 0	1, 2	0, 0
BED	0, 0	0, 0	10, 10

Figure 3 depicts the battle-of-the-sexes-with-an-inside-option game. The upper-left 2×2 portion of the game is a standard battle-of-the-sexes game. In addition, each player has a third strategy, corresponding to staying home in bed. With the inside option, the game becomes a game of common interest, but not a game of equilibrium common interest because players have different rankings over the inefficient equilibria (BALLET, BALLET) and (FIGHT, FIGHT).

The game has three strict equilibria, which as singletons are the only EES sets. Consider a communication game in which there are two signals, 1 and 2. For this game, the strategy in which players choose each signal with probability 1/2, and then go to the BALLET if the sum of the indices of the signals is odd and go to the FIGHT otherwise. This strategy profile, as a singleton, is an EES set. To see this, observe that since each player uses all of its signals with positive probability, and then chooses strict equilibrium actions in the underlying game, any invading strategy must agree with the population strategy given any pair of signals. Hence the communication game reduces to a constant-sum game in which the unique equilibrium strategy for each player is to randomize equally over both signals. No other choice of signaling strategy could satisfy (iii). In this example, communication does not force the players to arrive at their favorite equilibrium. Instead players waste all of their words arguing over which inefficient equilibrium to play.

One method to avoid arguments is to assume that the message space has many elements or that there are many rounds of communication, and assume that each individual in the population uses a pure signaling strategy. In section Ⅸ we discuss this approach.

Another possibility is to strengthen the stability notion by requiring the population strategy to resist invasions from groups that are not arbitrarily small. If the cardinality of M is sufficiently great, then there will always be a pair of signals that are used with arbitrarily small probability in any equilibrium. A strategy that uses these signals to coordinate on the efficient equilibrium will not be an optimal response to the population if the invaders make up a sufficiently small portion of the population, but if invaders occur in lumps, so that there is a positive lower bound to the ε in (iii), then there always exists a large enough strategy space so that invasion is possible when the payoff is not efficient.

In section Ⅶ we modify the stability notion in order to rule out arguments.

Ⅶ. COMMUNICATION STABILITY

We demonstrated in the previous section that pre-play communication need not lead to efficiency in our model. The message space can be jammed with the players arguing over which of several equilibria to play. The argument must be conducted over equilibria that are not Pareto ranked, but all of the equilibria under discussion could be inefficient in the set of equilibria. It is in precisely this type of situation that one would expect new messages to be invented and used to reach the efficient outcome. In this section we introduce a more restrictive notion of stability that requires outcomes to persist even if new messages are permitted, and prove efficiency and existence theorems. The central insight is that by allowing players to create new messages (at minimal cost) and applying the appropriate stability concept, we are able to obtain versions of the existence and efficiency theorems of Section V without making the restrictive assumptions of common interests or equilibrium common interests.

Our goal is to characterize the EES sets of the communication game with message space M (the M game) that remain stable when additional messages are added to the language. When N strictly contains M, a strategy for the M game is not a strategy for the N game: If the set of messages expands, then the strategy must include a specification of whether to use the new messages and how to interpret them if the opponent uses them. Therefore, it does not make sense to require that an element of a stable set of the M game is also an element of a stable set of the N game. Nevertheless, the additional words in N can be superfluous; by extending strategies from the M game to the N game we can find out when additional words would upset an equilibrium.

There is a natural way to extend a strategy of the M game to make it a strategy of the N game. Let f: $N \rightarrow M$ be a function with the property that for all $m \in M$, $f(m) = m$. Given a strategy $\sigma = (\mu, \tau)$ of the M game, define the **extension** $\sigma^N = (\mu^N, \tau^N)$ to the N game by

$$\mu^N(n) = \begin{cases} \mu(n) & \text{if } n \in M \\ 0 & \text{if } n \notin M \end{cases} \quad \text{and} \quad \tau^N(t; n_1, n_2) = \tau(t; f(n_1), f(n_2)).$$

The definition of $\mu^N(\cdot)$ states that the extension of σ signals in precisely the same way as σ; in particular, the new messages are not sent. The extension responds to messages from M exactly as the original strategy. In addition, it interprets a new word as if it is one of the old ones. The definition of $\tau^N(\cdot)$ asserts that $f(\cdot)$ acts as a translator: it turns words from the larger language N into words of M. If new words are really superfluous, then they need not be used, and, if used, can be interpreted as existing words. Our notion of communication stability requires that if σ is an element of an EES set Θ in the M game, then any strategy that can invade when the population plays σ^N does not lead to a short-term gain to the invader.

Definition. An EES set for Θ for a communication game (M, T, U) is **communication stable** if for all $N \supset M$ and $\sigma \in \Theta$, if there exists $\varepsilon' > 0$ such that for all $\varepsilon \in (0, \varepsilon')$, $C(\sigma') \subset BR(\varepsilon\sigma' + (1-\varepsilon)\sigma^N)$, then for $i = 1$ and 2, $j \neq i$, and $\varepsilon \in (0, \varepsilon')$,

$$U_i(\sigma_i', \varepsilon\sigma_j' + (1-\varepsilon)\sigma_j^N) \leq U_i(\sigma_i^N, \varepsilon\sigma_j' + (1-\varepsilon)\sigma_j^N).^{6} \tag{10}$$

$C(\sigma') \subset BR(\varepsilon\sigma' + (1-\varepsilon)\sigma^N)$ implies that σ' is an optimal response to the perturbed population strategy $\varepsilon\sigma' + (1-\varepsilon)\sigma^N$. If we did not assume that $C(\sigma') \subset BR(\varepsilon\sigma' + (1-\varepsilon)\sigma^N)$, then it would be possible for strategies to enter and use additional messages to coordinate on a nonequilibrium outcome in the underlying game. It follows from $C(\sigma') \subset BR(\varepsilon\sigma' + (1-\varepsilon)\sigma^N)$ that the inequality in (10) can be taken to be an equation.

Condition (10) states that an invading strategy σ' can only enter the population if it is an optimal response to the population strategy and it performs better against itself than the population strategy performs against it $(U_i(\sigma_i', \sigma_j') > U_i(\sigma_i^N, \sigma_j'))$. Communication stability is therefore similar to the idea of a neutrally stable strategy.[7] It does not permit a strategy that uses new messages to grow unless that strategy is an optimal response to the perturbed (post-entry) environment and the population strategy is not an optimal response to the perturbed environment.

Communication stability is the correct criterion to apply if coining a new

---

[6] In this section $U_i(\sigma_i, \sigma_j)$ denotes the payoff to player i if player i uses σ_i and player $j \neq i$ uses σ_j.

[7] To be a neutral ESS a strategy σ must have the property that no invading strategy can do strictly better than it when matched with a population that contains a small fraction of individuals playing the invading strategy (and the rest playing σ).

word requires paying a cost that is infinitesimal relative to the payoffs in the underlying game. If the population is playing an efficient Nash equilibrium strategy, then there is no incentive to create new words. If there is an equilibrium payoff that dominates the EES payoff, then it is worthwhile for players to try to talk their way to a better outcome.

Extensions are well defined given $f(\cdot)$. For what follows, the choice of $f(\cdot)$ does not matter: A set is communication stable with respect to extensions defined by one $f(\cdot)$ if and only if it is communication stable with respect to extensions defined by all $f(\cdot)$ that satisfy $f(m)=m$ for all $m \in M$.

We view communication stability as a formalization of the rich-language assumptions that appear in the work of Farrell [1993], Matthews, Okuno-Fujiwara, and Postlewaite [1991], and Rabin [1990]. These authors assume that there always exist unused messages that have, in contrast to our approach, natural interpretations that all players understand and will believe if they do not conflict with strategic considerations. Farrell [1993], Matthews, Okuno-Fujiwara, and Postlewaite [1991], and Rabin [1990] study games with incomplete information and their results are not comparable to ours.

Proposition 4, which we state and prove below, demonstrates that the requirement of communication stability rules out inefficient payoffs. We prove the proposition in the appendix. For intuition, consider again the inefficient EES set in the battle-of-the-sexes-with-an-inside-option game discussed in Section VI. If the language M expands to include another word, the extension of the equilibrium will not be in an EES set. The extra word will be used by invaders to coordinate on the efficient equilibrium. This property is true even when there are multiple efficient equilibria.

Let Π denote the convex hull of Nash equilibrium payoffs of the underlying game.

Proposition 4. If Θ is a communication stable EES set of a game with pre-play communication, then $U(\sigma)$ is an undominated element in Π for all $\sigma \in \Theta$.

All of our results suggest that the only possible stable outcome must induce an efficient Nash equilibrium in the underlying game. We cannot expect stable sets to exist unless the efficient Nash equilibria of the underlying game are elements of stable sets themselves. Essentially this is the only assumption needed in order to guarantee that communication stable EES sets exist.

Definition. $\pi^* \in \Pi$ is a **regular efficient payoff** if $\pi \in \Pi$ and $\pi \geq \pi^*$ implies $\pi = \pi^*$ and there exist EES sets of (T, u) with payoffs u^* and v^*, and $\lambda \in [0,1]$ such that $\pi^* = \lambda u^* + (1-\lambda)v^*$.

If $\pi^* \in \Pi$ is undominated, then it is a convex combination of at most two equilibrium payoffs of the underlying game. A regular efficient payoff can be attained by averaging EES payoffs. Without this condition π^* cannot be an EES payoff in the communication game. All efficient payoffs will be regular if, for example, all of the efficient Nash equilibria in (T, u) are strict.

Proposition 5. Let $\pi^* \in \Pi$ be a regular efficient payoff. Given any $\delta > 0$, there exists a message space M and a communication stable EES set of (M, T, U) that leads to payoffs within δ of π^*.

Proof. First suppose that π^* is the payoff of a Nash equilibrium in the underlying game and that there exists a stable set Θ of (T, u) that leads to the payoff π^*. Let M contain only one message. Θ gives rise to a stable set of the communication game.

Now suppose that π^* is not an equilibrium payoff of the underlying game but that $\pi^* = \lambda u(r^*) + (1-\lambda) u(s^*)$, where r^* and $s^* \in N(T, u)$ (hence $u(r^*) \neq u(s^*)$ and $\lambda \in (0,1)$), and there exist stable sets Θ_r and Θ_s of (T, u) such that $r^* \in \Theta_r$ and $s^* \in \Theta_s$.

Find relatively prime positive integers k and l such that $l > k$ and

$$\| \{ [k/l] u(r^*) + [(l-k)/l] u(s^*) \}, \pi^* \| < \delta,$$

where if x and $y \in R^2$, then $\| x, y \|$ is Euclidean distance between x and y. Let M consist of the first l integers and Θ consist of all stategies $\sigma = (\mu, \tau)$ such that $\mu_i(m) = 1/l$ for $m \in M$ and $i = 1$ and 2, and

$$\tau(\cdot\,; m_1, m_2) \in \Theta_r \text{ if } m_1 + m_2 < k \pmod{l} \text{ and} \tag{11}$$

$$\tau(\cdot\,; m_1, m_2) \in \Theta_s \text{ if } m_1 + m_2 \geq k \pmod{l}, \tag{12}$$

where for positive integers k and l, $k \pmod{l}$ is the remainder when k is divided by l. First we verify that Θ is an EES set. Since Θ_r and Θ_s are EES sets, it must be the case that if $\sigma' = (\mu', \tau')$ can invade Θ, then it specifies actions that satisfy (11) and (12). Consequently, $C(\sigma') \subset BR((1-\varepsilon)\sigma + \varepsilon\sigma')$ for $\varepsilon > 0$ if and only if $C(\sigma') \subset BR(\sigma')$. Since k and l are relatively prime, $C(\sigma') \subset BR(\sigma')$ only if $\mu'(\cdot) = \mu(\cdot)$.

It remains to show that these sets are communication stable. Enlarge the message space to N and fix a translation function f. Let $\sigma \in \Theta$ and let $\sigma' = (\mu', \tau')$ satisfy (iii). It follows from $C(\sigma') \subset BR((1-\varepsilon)\sigma^N + \varepsilon\sigma')$ that (σ_i', σ_j^N) and (σ_i', σ_j') (for $i \neq j$) induce Nash equilibria on the underlying game following any pair of messages, and that the equilibria are elements of either

Θ_r or Θ_s when at least one of the messages is an element of M (which guarantees that both messages are interpreted as messages from M using f). Therefore,

$$U_i(\sigma_i^N, \sigma_j') = [k/l]u_i(r^*) + [(l-k)/l]u_i(s^*) \tag{13}$$

because, for example, no matter what message player one sends, (σ_1', σ_2^N) and (σ_1^N, σ_2') must induce the same mixture between equilibria in Θ_r and Θ_s. Since σ_i' is an optimal response to σ_j^N,

$$U_i(\sigma_i', \sigma_j^N) \geq U_i(\sigma_i^N, \sigma_j^N) = [k/l]u_i(r^*) + [(l-k)/l]u_i(s^*). \tag{14}$$

Finally, observe that $U(\sigma_1', \sigma_2^N)$ and $U(\sigma_1^{N\prime}, \sigma_2')$ must be elements of Π, and that $[k/l]u(r^*) + [(l-k)/l]u(s^*)$ is an efficient element of Π. Consequently (10) follows from (13) and (14).

There are two unsatisfactory aspects to the existence theorem. First, there may be many efficient Nash equilibrium payoffs in the underlying game, some of which are not EES payoffs. This does not mean that the others should be ignored. It means that a dynamic process that describes the invasion need not settle down. Even if one thinks that the stability conditions captured by EES and communication stability are appropriate, Propositions 4 and 5 only demonstrate that processes describing these stability conditions could settle down to only a subset of feasible payoffs for the underlying game, and that subset is contained in the convex hull of the set of equilibrium payoffs. Dynamic behavior that does not settle on a particular payoff seems to be a likely outcome of a fully specified dynamic adjustment process.

The second difficulty is that we must fix the size of the message space in advance. This restriction is strong when there is a unique efficient equilibrium payoff in the underlying game. The game in Figure 2 demonstrates that EES sets need not exist in a communication game (M, T, U) when there is a unique efficient equilibrium payoff in (T, u) and M has cardinality greater than one. For general games it is communication stability, a restriction imposed in order to model the possibility of communication, that forces efficiency. Allowing pre-play communication only enables players to obtain mixtures of Nash equilibria. When there are multiple efficient stable equilibria in the underlying game, the restriction is less severe: Any regular efficient payoff can be approximated by payoffs from an EES set for a communication game with an arbitrarily large message space.

Ⅷ. ONE-SIDED COMMUNICATION

Thus far we have assumed that both players had access to languages with the same number of words. If both players have access to potentially different, but arbitrarily rich, languages, then Propositions 1 through 4 continue to hold. Since Proposition 5 shows that it is not necessary to create asymmetric languages to guarantee that a family of communication stable EES sets exists, allowing players to have different message spaces does not allow us to strengthen the result. If one or both of the players have access to bounded languages, then not all of the mixtures of equilibria will be feasible. At the least, the notion of efficiency used in Section Ⅶ must be modified to take into account that the set of Nash equilibrium payoffs of communication games would not be dense in the convex hull of Nash equilibrium payoffs of the underlying game. We do not pursue this restriction. Instead, in this section we consider the case in which only one player is able to speak.

[Figure 4]

0, 0	13, 5	5, 13
0, 1	5, 13	13, 5
4, 4	7.5, 1	8.5, 1
6, 0	6, 0	6, 0

Consider the game in Figure 4, which is taken (with a tiny modification) from Swinkles [1992]. There are three Nash equilibrium outcomes in the game. In the first, the strategies are ((1/2,1/2,0,0), (0,1/2,1/2)), which yields payoffs (9,9). In the second, the row uses (0,0,0,1) and the column player chooses a strategy to which this is an optimal response; the payoff of this equilibrium is (6,0). In the third the strategies are ((3/23,3/23,17/23,0), (1/5,2/5,2/5)) and the payoffs are (36/5,71/23). Since the payoffs associated with the different outcomes are Pareto ranked, this is a game of equilibrium common interest. The first equilibrium as a singleton is an EES set. The second set of strategies is an EES set, while the third equilibrium fails to be a part of an EES set (players coordinating on the first equilibrium can invade). Allowing only the row player to talk in this game does not rule out the outcome in which the row player chooses its fourth strategy in the underlying game no matter what it says. The stable set that supports this behavior in the communication game contains all signaling strategies for the row player provided that it plays the fourth strategy in the underlying game and requires that the column player play a strategy that supports the equilibrium independent of what the row player says. A strategy in which the row player uses one of its first three strategies in the underlying game cannot invade for the same reason that it could not invade when there is no communication: A mixture of row's first two strategies is never a best response to a strategy of column that supports the equilibrium.[8]

The example brings out a feature of our approach to pre-play communication that typically does not appear when language is assumed to have a focal meaning. In that literature, if only one player is able to communicate, then that player is assumed to be guaranteed to achieve its favorite equilibrium.

As Thoreau observed, effective communication requires two parties: one to speak and another to understand. When both players have rich languages, they are able to use their language to demonstrate their ability to understand. If only one player can speak, then attempts to communicate need not succeed.

Imagine that most of the population of column players is deaf to the new word. If the word GRYNSPAN is only used by communicators, then players who use the language to reach efficient outcomes can do so at no cost; they are able to modify their behavior accordingly if matched with players who do not communicate. When there is only one-sided communication the speaker does not know whether its opponent is listening. If the population consists primarily of agents who do not listen, then a player who wishes to use a new word to signal its intentions must plan to use a strategy that not only leads to a good equilibrium, but also responds optimally to the deaf population. The game in Figure 3 demonstrates that it is not always possible to do this.

One-sided communication leads to efficiency in some circumstances. For example in the game in Figure 1 there are two Pareto-ranked singleton EES sets. Without communication there is no way to guarantee that the players will arrive at the efficient outcome. Nevertheless, any potential EES set that leads to the payoffs (1,1) must contain the strategy in which the row player always says one thing (ALAS) and plays BAD and the column player responds by playing BAD, but if the row player said something else (GO FOR IT), the column player plays (1/2,1/2). A strategy profile in which the row player says GO FOR IT and plays GOOD and the column player plays GOOD given the other signal (and BAD otherwise) can invade. Similar reasoning demonstrates that the player who talks can get its favorite outcome in the battle-of-the-sexes game. In fact, in any 2×2 game, if one player can talk, then the only possible EES payoff for that player is its most preferred equilibrium payoff.

While one-sided communication does not guarantee efficiency, the talker's favorite equilibrium payoff will be communication stable in some game with

[8] Formaly, if an invading strategy of the row player is going to use one of the first three strategies with positive probability, then it must be that column's response is in the set D= $\{(L,C,R) : L,C,R \geq 0,\ L+C+R=1,\ 13C+5R \leq 6,\ 5C+13R \leq 6,$ and $4L+7.5C+8.5R \geq 6$ with equality in at least one of the three inequalities$\}$. It is straightforward to check that no strategy profile in which the row player plays one of its three strategies with positive probability can satisfy condition (iii) in the definition of EES for any response in D.

one-sided pre-play communication provided that the payoff is a regular efficient payoff (that is, it can be attained in an EES set of the underlying game). However, even if only the row player can talk in the nonexistence example in Section V (Figure 2), the arguments there demonstrate that no EES set exists when the row player has more than one message available.

Ⅸ. CONSECUTIVE COMMUNICATION

This section discusses communication in which players speak consecutively rather than simultaneously. The framework differs from the previous section in that we allow both players an opportunity to communicate. Without loss of generality we assume that there are just two rounds of communication. In the first round player one (the Row player) sends a message. Player two (the Column player) hears this message and sends a message of her own in the second round. Finally the players choose actions for the underlying game contingent on both messages. In this framework, we show that arguments cannot occur: Any EES set must have an efficient payoff.

Let u^* be an efficient point of the set of Nash equilibrium payoffs of (T, u).

Proposition 6. Assume M_i has at least two elements for $i=1$ and 2. If Θ is an EES set of (M, T, U), then there exists $\sigma \in \Theta$ such that $U(\sigma) = u^*$.

The method of proof of Proposition 6 is very similar to those of Propositions 2 and 3 and it consists of four parts (The detailed proof is provided in Kim and Sobel [1992]). The first two steps of the argument demonstrate that every stable set contains a strategy in which the players send only one of the messages with positive probability. If one player is randomizing over several strategies, there is nothing to stop the population from drifting to a situation in which he uses only one of his messages. The second step extends the first step and shows that each stable set must have an equilibrium in which both players use only one word. If the population of first players is using only one word, then the population of second players can drift to a situation in which it is using one of the words that leads to the highest payoff. The last two steps are nearly identical to the arguments in Section V. The third step shows that the population can drift to a strategy in which each side uses only one word, but that different messages do not lead to lower payoffs, and the fourth step demonstrates that from such a configuration a strategy that can obtain an efficient payoff is able to invade.

The main difference between this result and Proposition 4 is that we guarantee that a stable set contains an efficient outcome without invoking communication stability: Arguments cannot occur if players are not allowed to speak at the same time. Allowing both players a chance to talk does eliminate the possibility of sticking at an inefficient equilibrium, because it gives the second player the opportunity to acknowledge that she has heard and understood that the first player has tried to move from an inefficient equilibrium (two-sided communication makes it possible to use an unsent message without being punished by the dominant population).

When players must speak at the same time, convex combinations of Nash equilibria of the underlying game can be stable in the communication game. Nontrivial forms of correlation are not possible when players speak sequentially because player two will always make a statement that leads to her highest expected payoff. Therefore no element of a stable set can induce a mixture of Nash equilibria to the underlying game that give different payoffs to the second player to speak.

Just as there is an existence problem when players send messages simultaneously, there is an existence problem when players take turns speaking in games without common interests. It is straightforward to show that if we use communication stability then versions of Propositions 4 and 5 apply. Specifically, each communication stable set in a game with alternating communication must contain an efficient payoff, and given any Nash equilibrium payoff of the underlying game that is regular and undominated by other Nash equilibria there exist message spaces for which the result is communication stable (generally the message space will be trivial for at least one player, however).

Ⅹ. RELATED WORK

In this section we discuss other papers that use the evolutionary approach to model communication in games.

Wärneryd [1991] characterizes neutrally stable strategies in pure-coordination games with pre-play communication and complete information. Players must use pure strategies. In 2×2 games, he obtains the efficiency result. In larger games the result does not hold: Neutrally stable strategies that support an inefficient equilibrium exist provided that the population is able to punish invading strategies by switching to an even less efficient equilibrium.

Bhaskar [1998] also examines complete information games preceded by a round of simultaneous signaling played by a single population; he permits randomization at the individual level. He shows that unless there is a countably infinite set of messages, neutrally stable strategies need not lead to efficiency even in pure-coordination games. We pointed out in Section Ⅲ that a similar result holds for our solution concept if we assume that a single population plays the game.

Bhaskar obtains Wärneryd's [1991] results (that neutrally stable strategies must be efficient in 2×2 common-interest games with pre-play communication, but need not be for larger games) when players make small mistakes in signaling.[9]

Bhaskar also studies a model in which players may misinterpret their opponent's message. This type of mistake rules out punishments for generic underlying games and permits Bhaskar to prove a version of Proposition 2.

Fudenberg and Maskin [1991] also study the effect of imposing evolutionary stability on games with pre-play communication. They assume that the underlying game is finite and symmetric, that there is a potentially unlimited number of rounds of pre-play communication in which the players speak simultaneously. They further assume that players make mistakes with small probability. Talk is not completely free, but its cost is infinitesimal relative to the probability of mistakes.

Fudenberg and Maskin obtain an efficiency result. Any evolutionarily stable payoff must give each player at least as much as it gets in its least favorite strongly efficient outcome (where an outcome is strongly efficient if it maximizes the sum of payoffs) in any evolutionarily stable payoff. This result is stronger than ours in two respects. First, outcomes are not stable in Fudenberg and Maskin's sense if they are inefficient relative to the set of feasible payoffs rather than relative to the (smaller) convex hull of the set of equilibria in the underlying game. This difference is the result of our restriction to equilibrium entrants. The prisoners dilemma with pre-play communication fails to have a weak ESS because players who used different messages could invade the population and then cooperate with each other. As we explained earlier, EES does not allow this type of invasion. The set of all equilibria in which individuals never cooperate is an EES set.

Second, arguments are not possible in the Fudenberg-Maskin model. The population consists of a finite number of individuals who play pure communication strategies. There always exists a communication history following which some message is not used. An appropriately chosen message of this form avoids arguments.

Fudenberg and Maskin are able to prove existence if there is a symmetric, strict, strongly efficient equilibrium in the underlying game. This condition holds automatically in symmetric games with common interests. The result would hold for a version of our solution concept that requires symmetric entry, but does not hold for the asymmetric version we use in the paper.[10] Fudenberg and Maskin can also prove existence if there exists a strict symmetric equilibrium that is better than a strict strongly efficient equilibrium. This result does not hold in

[9] Unlike Wärneryd's [1991], Bhaskar [1998] must assume that players tremble in order to obtain the efficiency result in the 2×2 case. Otherwise one can support an inefficient outcome with a randomized punishment.

[10] Consider the game in Figure 5 with w=4. When we allow pre-play communication (using a nontrivial language), there is no EES set that yields payoff three to each player because some element of a set supporting this payoff would permit the entry of a group of invaders to recognize each other and coordinate on one of the asymmetric strict equilibria when they play each other while playing the symmetric strict equilibrium against the original population.

our setting, as demonstrated by the variation of the battle-of-the-sexes game in Figure 5 (with w>5). The strict symmetric equilibrium in which all players choose their SAFE strategy is evolutionarily stable. Because players have identified roles and are able to use mixed signaling strategies, they are able to correlate on an outcome that alternates between (BALLET, FIGHT) and (FIGHT, BALLET) with equal probabilities. This correlated equilibrium dominates the outcome in which individuals always play SAFE. Consequently, an outcome that yields payoff three will not be communication stable. It will also fail to be part of an EES set when the language is nontrivial.

[Figure 5]

	BALLET	FIGHT	SAFE
BALLET	0, 0	w, 1	0, 0
FIGHT	1, w	0, 0	0, 0
SAFE	0, 0	0, 0	3, 3

Matsui [1991] applies a variation of the Gilboa and Matsui [1991] idea of cyclically stable sets to show that the only cyclically stable set in 2×2 common-interest games with pre-play communication contains only efficient equilibria. The corollary of Section 5 contains the conclusions of this theorem. The essential difference between cyclically stable sets (CSS) and EES sets is that cyclically stable sets of strategies need not satisfy condition (ii) of EES.[11] The efficiency result does not generalize to larger games because arguments can arise (the inefficient strategy that is an EES set for the example in Section V is also a CSS). Also, since Matsui does not impose a requirement that stable strategy profiles be equilibria, CSSs consisting only of inefficient, nonequilibrium strategies will exist for 3×3 (or larger) games.

Sobel [1993] presents existence and efficiency results for common-interest games with pre-play communication using a static evolutionary stability concept that, like Gilboa and Matsui's cyclically stable sets, does not assume equilibrium behavior. He obtains an efficiency result for general common-interest games assuming a finite population of players who use only pure strategies. The paper uses its stability concept to obtain efficiency results for two different types of common-interest games, infinitely repeated games and incomplete-information games with cheap-talk. It contains a survey of other papers that apply evolutionary stability to these games.

No one has presented a dynamic foundation for the EES concept. There is the question of whether there is any dynamic foundation to the results that we present in this paper. There is a connection between local stability of an

[11] Matsui [1992] compares EES sets to cyclically stable sets.

outcome under the replicator dynamic (or closely related processes) and outcomes that satisfy static stability conditions. In particular, Cressman [1992] shows that in symmetric games Thomas's ES sets are necessarily locally stable for pure-strategy dynamics. In common-interest games, the efficient outcomes that we identify as evolutionarily stable, and only those outcomes, are locally stable with respect to the replicator dynamic.

Matsui and Rob [1991], Nöldeke, Samuelson, and van Damme [1991], and Kim and Sobel [1995] have shown that only efficient outcomes arise as limits of an evolutionary dynamic process in pure-coordination games with pre-play communication. These papers assume that the population of players is finite; that players change their strategies randomly; and that mistakes or mutations occur and cause the models to have a unique ergodic distribution, which they can characterize. As in this paper, movements arise because with positive probability the population will select an optimal response to its current configuration. The dynamics of these models exhibit the drift that occurs in EES sets. In fact, these dynamics permit even more drift: EES does not allow strategies to drift outside an equilibrium component, while in Nöldeke, Samuelson, and van Damme [1991] strategies can drift arbitrarily at unreached information sets.

APPENDIX

Proof of the lemma. Fix a strategy profile for the communication game $\sigma=(\mu,\tau)$. Let $(\widetilde{m}_1,\widetilde{m}_2)\in M\times M$ be a pair of messages such that $\mu_i(\widetilde{m}_i)>0$ for $i=1$ and 2. Now define the strategy profile $\sigma'=(\mu',\tau')$ such that $\mu'(\cdot)=\mu(\cdot)$,

$$\tau_1'(t_1;m_1,m_2)=\begin{cases}\tau_1(t_1;m_1,m_2) & \text{if } m_2\neq\overline{m_2}\\ \tau_1(t_1;m_1,\widetilde{m}_2) & \text{if } m_2=\overline{m_2}\end{cases}\quad\text{and}$$

$$\tau_2'(t_2;m_1,m_2)=\begin{cases}\tau_2(t_2;m_1,m_2) & \text{if } m_1\neq\overline{m_1}\\ \tau_2(t_2;\widetilde{m}_1,m_2) & \text{if } m_1=\overline{m_1}\,.\end{cases}$$

Since $\sigma\in N(M,T,U)$, σ' differs from σ only following $(\overline{m_1},\overline{m_2})$, and players can do no better using $\overline{m_i}$ than other messages, it follows that $\sigma'\in N(M,T,U)$. Furthermore, $C(\sigma')\in BR(\sigma)$ because σ is a Nash equilibrium and σ' agrees with σ on the equilibrium path. Consequently the lemma follows from Proposition 1.

Proof of Proposition 2. First we show that there exists $\sigma\in\Theta$ that does not use all signals. Let $\hat{\sigma}=(\hat{\mu},\hat{\tau})\in\Theta$, and let $(\widehat{m}_1,\widehat{m}_2)$ lead to the highest payoff for player one of all message pairs sent with positive probability under $\hat{\sigma}$. Since the underlying game has ECI and $\hat{\sigma}$ is a Nash equilibrium, this message pair also leads to player two's highest payoff. Define $\sigma=(\mu,\tau)$ so that $\tau(\cdot)\equiv\hat{\tau}(\cdot)$ and $\mu_i(m_i)=\begin{cases}0 & \text{if } m_i\neq\widehat{m}_i\\ 1 & \text{if } m_i=\widehat{m}_i\end{cases}$. σ_i is an optimal response to both σ_j and $\hat{\sigma}_j$ for $j\neq i$. Hence, by Proposition 1, $\sigma\in\Theta$.

Take $\sigma=(\mu,\tau)\in\theta$ as above (so that $\mu_i(\cdot)$ places probability one on a proper subset of M) and assume, in order to obtain a contradiction, that $U(\sigma)\neq u^*$. Let $\sigma'=(\mu',\tau')$ satisfy the conclusion of the lemma. Let s^* be a Nash equilibrium strategy profile of the underlying game such that $u(s^*)>U(\sigma')$. Define $\sigma''=(\mu'',\tau'')$ where

$$\mu_i''(m_i)=\begin{cases}0 & \text{if } m_i\neq\overline{m}_i\\ 1 & \text{if } m_i=\overline{m}_i\end{cases}\quad\text{and}$$

$$\tau''(t;m_1,m_2)=\begin{cases}\tau'(t;m_1,m_2) & \text{if } (m_1,m_2)\neq(\overline{m}_1,\overline{m}_2)\\ s^*(t) & \text{if } (m_1,m_2)=(\overline{m}_1,\overline{m}_2)\end{cases}.$$

Since $s^*\in N(T,u)$ and $u(s^*)\geq U(\sigma)$, $\sigma''\in N(M,T,U)$. Furthermore, $C(\sigma'')$

$\subset BR(\sigma')$ by the lemma. It follows from Proposition 1 that $(1-\varepsilon)\sigma' + \varepsilon\sigma'' \in \Theta$ for all $\varepsilon \in [0,1]$ and hence, by the definition of EES sets, $(1-\varepsilon)\sigma' + \varepsilon\sigma'' \in N(T,u)$ for all $\varepsilon \in [0,1]$. This leads to a contradiction since, for $\varepsilon \in (0,1)$, $(1-\varepsilon)\sigma' + \varepsilon\sigma'' \notin N(T,u)$ (the unique optimal message for player one, given that the other player uses $(1-\varepsilon)\sigma_2' + \varepsilon\sigma_2''$ is $\overline{m}_1$, but under $(1-\varepsilon)\sigma_1' + \varepsilon\sigma_1''$ player one uses other messages with positive probability).

Proof of Proposition 4. Take $\sigma \in \Theta$ and assume, in order to reach a contradiction, that $U(\sigma)$ is a dominated element of Π. By the lemma, we may also assume without loss of generality that for each $m \in M$, there exists $(m, t_i) \in BR_i(\sigma_j)$ for $j \neq i$. Since Π is a nonempty, compact, convex subset of R^2, there exist a pair of Nash equilibrium strategies for the underlying game, call them r^* and s^* (r^* could be equal to s^*), such that $u(r^*)$ and $u(s^*)$ are extreme points of Π, and positive integers k and l such that $l \geq k$ and

$$[k/l]u(r^*) + [(l-k)/l]u(s^*) > U(\sigma). \tag{A2}$$

Construct an N game in which N has l messages in addition to those in M. Denote the additional messages by 1, 2,$\cdots$, l. Consider the strategy $\hat{\sigma} = (\hat{\mu}, \hat{\tau})$ defined by

$$\hat{\mu}_i(m_i) = \begin{cases} 0 & \text{if } m_i \in M \\ 1/l & \text{if } m_i \notin M \end{cases} \quad \text{and}$$

$$\hat{\tau}(t; m_1, m_2) = \begin{cases} \tau_N(t; m_1, m_2) & \text{if } m_1 \text{ or } m_2 \in M \\ r^*(t) & \text{if } m_1 + m_2 < k \ (\mathrm{mod}\, l) \text{ and} \\ & \quad m_i \notin M \text{ for } i = 1 \text{ and } 2 \\ s^*(t) & \text{if } m_1 + m_2 \geq k \ (\mathrm{mod}\, l) \text{ and} \\ & \quad m_i \notin M \text{ for } i = 1 \text{ and } 2. \end{cases}$$

It suffices to show that $C(\hat{\tau}) \subset BR(\varepsilon\hat{\sigma} + (1-\varepsilon)\sigma^N)$, but that (10) does not hold.

$\hat{\sigma}$ specifies that players use Nash equilibrium strategies following every pair of messages. Furthermore, given that one player signals according to $\hat{\mu}_i(\cdot)$, the other player (j) obtains the expected payoff $[k/l]u_j(r^*) + [(l-k)/l]u_j(s^*)$ if it uses any signal in $N \backslash M$, and expects no more than $U_j(\sigma)$ otherwise. It follows from (A2) that $\hat{\sigma}$ is a Nash equilibrium. $\hat{\sigma}$ is also an optimal response to the population strategy σ^N: Given that the other player is using σ_i^N, a player is indifferent between all of the messages in M, but messages in N are all treated as words in M by σ^N, so any signaling strategy is part of an optimal response to σ^N. Consequently, $C(\hat{\sigma}) \subset BR(\varepsilon\hat{\sigma} + (1-\varepsilon)\sigma^N)$.

When an agent playing the new strategy meets an agent playing the population strategy, both players obtain the payoff that they would have under the population strategy. If two players using the new strategy meet, then they obtain the payoff given by the left-hand side of (A2). It follows that

$$U_i(\hat{\sigma}_i,\ \varepsilon\hat{\sigma} + (1-\varepsilon)\sigma^N) = \varepsilon\{[k/l]u_i(r^*) + [(l-k)/l]u_i(s^*)\} + (1-\varepsilon) \quad \text{(A3)}$$

$U_i(\sigma^N)$ and

$$U_i(\sigma_i^N, \varepsilon\sigma_j + (1-\varepsilon)\sigma_j^N) = U(\sigma^N) = U(\sigma). \quad \text{(A4)}$$

Combining (A2), (A3), and (A4), we see that $U_i(\varepsilon\hat{\sigma} + (1-\varepsilon)\sigma^N) > U_i(\sigma_i^N, \varepsilon\hat{\sigma}_j + (1-\varepsilon)\sigma_j^N)$. Hence $\hat{\sigma}$ does not satisfy (10) and Θ cannot be communication stable.

REFERENCES

Bhaskar, V. (1998), "Noisy Communication and the Evolution of Cooperation," *Journal of Economic Theory*, 82, 110-31

Binmore, K. (1992), "Foundation of Game Theory," in J. J. Laffont, ed., *Advances in Economic Theory: Sixth World Congress*, Vol. 1, Cambridge University Press.

Cressman, R. (1992), "Evolutionarily Stable Sets in Symmetric Extensive Two-Person Games," *Mathematical Biosciences,* 108, 179-201.

Farrell, J. (1988), "Communication, Coordination, and Nash equilibrium," *Economics Letters*, 27, 209-14.

Farrell, J. (1993), "Meaning and Credibility in Cheap-Talk Games," *Games and Economic Behavior,* 5, 514-31.

Foster, D. and P. Young (1990), "Stochastic Evolutionary Game Dynamics," *Theoretical Population Biology*, 38, 219-32.

Fudenberg, D. and E. Maskin (1991), "Evolution and Communication in Games," Harvard University, preliminary notes.

Gilboa, I. and A. Matsui (1991), "Social Stability and Equilibrium," *Econometrica*, 59, 859-67.

Kandori, M., G. Mailath, and R. Rob (1993), "Learning, Mutation, and Long Run Equilibria in Games," *Econometrica*, 61, 29-56.

Kim,Y.-G. and J. Sobel (1992), "An Evolutionary Approach to Pre-Play Communication," UCSD Discussion Paper.

Kim,Y.-G. and J. Sobel (1995), "An Evolutionary Approach to Pre-Play Communication," *Econometrica*, 63, 1181-94.

Kohlberg, E. and J.-F. Mertens (1986), "On the Strategic Stability of Equilibria," *Econometrica*, 54, 1003-38.

Matsui, A. (1991), "Cheap-Talk and Cooperation in a Society," *Journal of Economic Theory*, 54, 245-58.

Matsui, A. (1992), "Best Response Dynamics and Socially Stable Strategy," *Journal of Economic Theory*, 57, 343-62.

Matsui, A. and R. Rob (1991), "The Role of Public Information and Pre-Play Communication in Evolutionary Games," University of Pennsylvania.

Matthews, S., M. Okuno-Fujiwara, and A. Postlewaite (1991), "Refining Cheap-Talk Equilibria," *Journal of Economic Theory*, 55, 247-73.

Maynard Smith, J. and G. Price (1973), "The Logic of Animal Conflict," *Nature*, 246, 15-8.

Myerson, R. (1983), "Mechanism Design by an Informed Principal," *Econometrica*, 51, 1767-97.

Myerson, R. (1989), "Credible Negotiation Statements and Coherent Plans," *Journal of Economic Theory*, 48, 264-91.

Nöldeke, G., L. Samuelson, and E. Van Damme (1991), "The Evolution of Communication," Universität Bonn, Discussion Paper.

Rabin, M. (1990), "Communication Between Rational Agents," *Journal of Economic Theory*, 51, 144-70.

Rabin, M. (1994), "Pre-Game Communication," *Journal of Economic Theory*, 63, 370-91.

Robson, A. (1990), "Efficiency in Evolutionary Games: Darwin, Nash and the Secret Handshake," *Journal of Theoretical Biology*, 144, 379-96.

Selten, R. (1983), "Evolutionary Stability in Extensive Two-Person Games," *Mathematical Social Sciences*, 5, 269-363.

Sobel, J. (1993), "Evolutionary Stability and Efficiency," *Economics Letters*, 42, 301-12.

Swinkels, J. (1992), "Evolutionary Stability with Equilibrium Entrants," *Journal of Economic Theory*, 57, 306-32.

Thomas, B. (1985a), "On Evolutionarily Stable Sets," *Journal of Mathematical Biology*, 22, 105-15.

Thomas, B. (1985b), "Evolutionarily Stable Sets in Mixed-Strategist Models," *Theoretical Population Biology*, 28, 332-41.

Van Damme, E. (1987), *Stability and Perfection of Nash Equilibria*. Berlin: Springer Verlag.

Wärneryd, K. (1991), "Evolutionary Stability in Unanimity Games with Cheap Talk," *Economics Letters*, 36, 375-8.

Wärneryd, K. (1993), "Cheap Talk, Coordination, and Evolutionary Stability," *Games and Economic Behavior*, 5, 532-46.