An Evolutionary Approach to Pre-Play Communication
Author(s): Yong-Gwan Kim and Joel Sobel
Source: *Econometrica,* Vol. 63, No. 5 (Sep., 1995), pp. 1181-1193
Published by: The Econometric Society
Stable URL: http://www.jstor.org/stable/2171726
Accessed: 10/12/2010 18:46

AN EVOLUTIONARY APPROACH
TO PRE-PLAY COMMUNICATION

By Yong-Gwan Kim and Joel Sobel[1]

We add a round of pre-play communication to a finite two-player game played by a population of players. Pre-play communication is cheap talk in the sense that it does not directly enter the payoffs. The paper characterizes the set of strategies that are stable with respect to a stochastic dynamic adaptive process. Periodically players have an opportunity to change their strategy with a strategy that is more successful against the current population. Any strategy that weakly improves upon the current poorest performer in the population enters with positive probability. When there is no conflict of interest between the players, only the efficient outcome is stable with respect to these dynamics. For general games the set of stable payoffs is typically large. Every efficient payoff recurs infinitely often.

KEYWORDS: Game theory, evolution, pre-play communication, cheap talk, adaptive behavior.

1. INTRODUCTION

ALTHOUGH INFORMAL STORIES in game theory emphasize that pre-play communication allows players to coordinate on efficient Nash equilibria, these stories 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 play 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), Myerson (1989), and Rabin (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 are stable with respect to a stochastic dynamic adaptive process. We assume that there is a finite population of players who use pure strategies. Periodically a player has an opportunity to replace his strategy by a more successful strategy. It is important for our results that the replacement process never ends (even when all strategies are performing equally well) and that all strategies which weakly improve on the poor performer are possible replacements.

The following example illustrates how pre-play communication enables the population to move away from inefficient equilibria. The coordination game in Figure 1 has two pure-strategy Nash equilibria and a completely mixed one. There is no conflict of interest in this game and one can supply many reasons why 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.

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. When players have more than one communication strategy before they reach the underlying game, the communication game will have no strict equilibrium; evolutionary pressures, which favor strategies that do well against the existing population, force efficiency.

In this paper we demonstrate that evolutionary stability leads to efficiency in two-player games where the interests of the players coincide in a strong way. In the example it is not difficult to see why outcomes that induce the BAD action are not stable. Suppose that the population has coordinated on an equilibrium in which the players always use a particular (normal) message. Everyone chooses the BAD action regardless of the message sent. As long as abnormal words are not used, however, there is no pressure to respond to them in a particular way. It is possible for an individual in the population to replace his strategy by one that sends the normal message, and plays GOOD in the underlying game if and only if he meets another player who sends an abnormal message. This strategy does as well as the original strategy and, over time, it is possible for the population to reach a state in which everyone would respond to an abnormal message with the GOOD action, although no one in the population actually uses the abnormal message. At this point a strategy that uses an abnormal message and then always plays GOOD does better than the existing population. This strategy can continue to replace any strategy that uses the BAD action. Consequently, it thrives. Furthermore, once the population has reached a configura-

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

FIGURE 1.

tion in which every pair of players is able to coordinate on the GOOD equilibrium, it is not possible to move away from this outcome. Eventually the population plays the efficient equilibrium and when it does there is no pressure to move to an inefficient outcome.

In the next section we describe the communication game and the adjustment process that we study. Section 3 demonstrates that, for games with common interests, where there is only one efficient payoff, cheap pre-play communication forces efficiency under our evolutionary dynamics. In Section 4 we describe an efficiency result for general games: Independent of what strategy the population plays initially, every efficient payoff recurs infinitely often. This result suggests that our dynamic process provides little guidance about how the game is played when the common-interest assumption fails. Section 5 extends the analysis to games in which only one player can speak. We show that if the speaker's preferred outcome is a strict equilibrium, then it is the only long-run outcome observed under our assumptions. Section 6 discusses related literature and variations of our model.

2. THE FRAMEWORK

We begin with a given finite two-player game,[2] which we call the underlying game. We add to the game one round of communication. There is a finite set M of messages (words) that contains at least two elements. For most of the paper we assume that players have access to the same set of messages and speak simultaneously. We discuss the case in which only some of the players can talk in Section 5. 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.

A finite population plays this communication game. Each member of the population is assigned to play either the role of player one or player two. These players use pure strategies. Pairs from the population play the game repeatedly and anonymously. On rare occasions, one member of the population changes his strategy. We look for stable sets of strategies for adaptive dynamics satisfying the properties described below.

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 com-

---

[2] 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.

munication 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 cannot ignore cheap talk.

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

$$(1) \qquad U_j(m_1, t_1(\cdot), m_2, t_2(\cdot)) = u_j(t_1(m_1, m_2), t_2(m_1, m_2)).$$

There are $2N$ individuals (N pairs) in the population. We will assume that players do not condition their strategies on the identity of their opponent. This assumption makes more sense when the population is large; our formal arguments require only that the population be finite. Denote the strategy of individual k in role i by $\theta_i(k) = (m_i, t_i(\cdot)) \in M \times T_i^{M \times M}$. A *population strategy profile* is a list that specifies a strategy $(\theta_1(n), \theta_2(n))$ for each agent $n = 1, \ldots, N$ and role $i = 1$ and 2 in the population. A strategy profile is *homogeneous* if $\theta_i(n)$ is independent of n for $i = 1$ and 2. Given a population strategy profile $\theta = (\theta_1(1), \theta_2(1), \ldots, \theta_1(N), \theta_2(N))$, the payoff functions U_1 and U_2 induce a population payoff for individual k_1 in role 1:

$$(2) \qquad \hat{U}_1(k_1; \theta) = \sum_n U_1(\theta_1(k_1), \theta_2(n))$$

and, for individual k_2 in role 2,

$$(3) \qquad \hat{U}_2(k_2; \theta) = \sum_n U_2(\theta_1(n), \theta_2(k_2)).$$

That is, we compute the population payoff for an individual under the assumption that he is matched exactly once with everyone assigned to the other role in the population.

Let $\theta / \theta_{i'}(k)$ denote the strategy profile with $\theta_i(k)$ replaced by $\theta_{i'}(k)$. We say that $\theta_{1'}(k)$ is a *best response* to the strategy profile θ if $\hat{U}_1(k; \theta / \theta_{1'}(k)) \geq \hat{U}_1(k; \theta / \theta_{1''}(k))$ for all pure strategies $\theta_{1''}(k)$. We say that $\theta_{1'}(k)$ *improves upon* $\theta_1(k)$ if $\hat{U}_1(k; \theta / \theta_{1'}(k)) \geq \hat{U}_1(k; \theta)$. Similar definitions hold for a player in the second role.

We assume that the population starts at an arbitrary strategy profile. To simplify analysis, we assume that in each round each player meets all of the other members of the population and plays the communication game once. Hence if the strategy profile at the beginning of a round is θ, then the payoffs from that round are given by (2) and (3). At the end of the round, one member of the population has the opportunity to change his strategy. In order to

describe the dynamic behavior of our system, we must specify the probability that any member of the population is allowed to change his strategy and identify the strategy that he changes to. These probabilities in principle could depend on the entire history of play. In what follows we assume that the probability that an individual is permitted to replace his strategy at the end of a round depends only on the current population strategy profile. The probability that he picks a particular replacement strategy depends only on the individual and the current population strategy profile. This stationarity assumption simplifies exposition, but is not necessary for our results. To handle more general specifications, substitute "with probability bounded away from zero" for "with positive probability" in Assumptions (S), (R), and (BR) below. We maintain four assumptions about the dynamic process.

ASSUMPTION (I): *Exactly one member of the population may change his strategy each round.*

ASSUMPTION (S): *Some individual who performs worst in his role in a round has the opportunity to change his strategy with positive probability.*

ASSUMPTION (R): *Any strategy that improves upon the strategy being replaced is adopted with positive probability.*

ASSUMPTION (NL): *A strategy that does not improve upon the agent's current strategy cannot be adopted.*

Assumption (I) states that adjustments are *individual*. This condition simplifies exposition. We use Assumption (I) to guarantee that efficient payoffs are stable in common-interest games. Assumption (S) is a *selection* condition. It requires that the evolutionary process replace poorly performing strategies. Assumption (R) requires that the replacement process be sufficiently *rich* that any strategy that weakly improves upon an existing strategy may be adopted. Under our evolutionary dynamics players may change to strategies that perform relatively well, although not necessarily optimally. The third assumption also requires that there be a positive probability that strategies not currently represented may replace existing strategies. This property would not hold for a deterministic biological dynamic (where any strategy that is not present in one generation cannot appear in subsequent generations). While (R) rules out a best-response dynamic that requires the replacement strategy be an optimal response to the current population, the results of Sections 3 and 5 would hold if we made the following weaker assumption:

ASSUMPTION (BR): *Any strategy that responds optimally to the population strategy is adopted with positive probability.*

Assumption (NL) assumes that agents can avoid strategies that perform poorly (under existing conditions); they will always pick new strategies that lead

to *no loss*. Qualitatively our results would not change if there is a probationary period after which strategies that are new to the population are replaced immediately if they fail to outperform their predecessor.

Let $A(\theta)$ be the set of strategy profiles that can be reached with positive probability starting from the profile θ. We extend this notation to sets of population profiles: For any set of profiles Θ, $A(\Theta)$ is the union of $A(\theta)$ for $\theta \in \Theta$. Θ is *absorbing* if $A(\Theta) \subseteq \Theta$. Call a minimal, nonempty, absorbing set *stable*. Since the set of all strategy profiles is absorbing and finite, stable sets exist. In fact, each absorbing set must contain a stable set. It is straightforward to check that

(4)   the intersection of absorbing sets is absorbing and

(5)   for each Θ, $A(\Theta)$ is absorbing.

By (4) two stable sets coincide or are disjoint. By (5) a stable set Θ^* satisfies $A(\Theta^*) = \Theta^*$ and also for each θ, there exists a stable set Θ^* such that $\Theta^* \subseteq A(\theta)$.

We call strategies that occur infinitely often with positive probability *recurrent*; other strategies are *transient*. For our model the recurrent profiles are precisely the elements of stable sets. A profile that is not an element of a stable set must be transient (since, by (R), whenever the profile arises it is eventually replaced with positive probability by an element of a stable set, and once the population profile is in a stable set it never leaves). A profile that is an element of a stable set must be recurrent (otherwise there is a smaller absorbing set contained in the stable set).

We use the structure of stable sets explained above to prove our main results. Given a population strategy profile θ, we identify elements of $A(\theta)$ by using properties (S) and (R). In particular, if a population strategy profile θ' is obtained from θ by making changes in how players respond to unsent messages, then $\theta' \in A(\theta)$. We refer to this type of change as *drift*.

3. GAMES WITH COMMON INTEREST

In this section we assume that the set of feasible payoffs $F = \{(u_1(t_1, t_2), u_2(t_1, t_2)): t_i \in T_i\}$ of the underlying game has a unique point (u_1^*, u_2^*) that strongly Pareto-dominates all other feasible payoffs (so that if $(u_1, u_2) \in F$, $(u_1, u_2) \neq (u_1^*, u_2^*)$ implies that $u_i < u_i^*$ for $i = 1$ and 2). In this case we say that the game has *common interests*. A game with common interests need not have a unique efficient action as several strategy combinations may give rise to the payoff (u_1^*, u_2^*). Denote by Θ^* the set of all population strategies that give rise to these payoffs (that is, the set of strategies θ such that $\hat{U}_i(k; \theta) = u_i^*$ for all $k = 1, \ldots, N$ and $i = 1$ and 2).

PROPOSITION 1: *In any game with common interests, if there are at least two messages, then the set Θ^* of all strategy profiles that give rise to efficient payoffs is stable.*

PROOF: It is sufficient to prove that $A(\theta) = \Theta^*$ for all $\theta \in \Theta^*$. First note that $\theta \in \Theta^*$ implies that $A(\theta) \subseteq \Theta^*$. Suppose that individual k in role 1 has an opportunity to replace his strategy. (NL) guarantees that the new strategy must lead to a payoff of u_1^*. Since (u_1^*, u_2^*) is the unique weakly efficient point of the game, each role two player in the population must obtain u_2^* when matched against individual k's new strategy.

It remains to show that $\theta \in \Theta^*$ implies $\Theta^* \subseteq A(\theta)$. Fix $\theta' \in \Theta^*$. It is possible for the population strategy to drift to a configuration in which all players in a given role use the same strategy (for example, if one by one individuals in role one adopt the same best response to the population strategy, and then individuals in the second role do the same). Call the messages used under this strategy (m_1, m_2). If these messages are not used in θ', responses to the other messages can drift so that they agree with θ', and a series of replacements can lead the population to θ'. Otherwise, the following sequence of replacements arises with positive probability: responses to another pair of messages drift to actions that support the efficient payoff; all individuals use these messages; the response to (m_1, m_2) drifts to the response played under θ'; the responses to the other messages drift so that they agree with θ'; and individuals replace their signaling strategies with the ones in θ'.

Proposition 1 requires that M, the set of messages, has at least two elements. If there are two different ways to achieve the efficient payoff in the underlying game, and M contains only one element, then there will be distinct stable sets supporting each equilibrium.

In order to guarantee that once the population arrives at an efficient payoff it is sure to stay there, we must use (NL). If inferior strategies enter with positive probability, then play could depart Θ^*. We would obtain the same qualitative result if we weaken (NL) but require that any new strategy that does not perform as well as the strategy that it replaces is immediately replaced. This dynamic behavior would be guaranteed if we assumed that only those individuals using the least successful strategy in the current population are replaced (a property that is consistent with (S)). Under these conditions once the population arrives at a strategy in Θ^* it stays close in that the population strategy profile differs from an element in Θ^* by the strategy of at most one individual.

Stability of Θ^* could also be destroyed if we weakened (I) to permit many individuals to change their strategies simultaneously.

Proposition 1 guarantees that the set of strategy profiles leading to the efficient payoff is a stable set for common-interest games. Next we show that there is a unique stable set of profiles, and that this set contains an efficient point. Combined with Proposition 1 it establishes that pre-play communication forces efficiency in common-interest games.

PROPOSITION 2: *In any game with common interests, if there are at least two messages, then there is a unique stable set.*

PROOF: For $i = 1$ and 2, let u_i^* be the highest feasible payoff for a player in role i of the underlying game; let (a_1, a_2) be strategies in the underlying game that give rise to a payoff (u_1^*, u_2^*). Let θ^* denote the homogeneous strategy profile in which each individual in the population sends the same message m and, independent of the messages sent, an individual in role i plays a_i. We will show that $\theta^* \in A(\theta)$ for all θ. It follows that θ^* is an element of any absorbing set; hence it is an element of any stable set. Since stable sets either coincide or are disjoint, there is a unique stable set, and it contains θ^*.

Through a series of replacements we can transform an arbitrary population profile θ first to a homogeneous profile in which unused messages lead to payoff u_1^*, then to a strategy profile in which role one players actually attain this payoff, and finally to θ^*. One by one, each agent in role one can be given an opportunity to replace his strategy with the same optimal response to the role two strategies currently in the population; suppose this leads agents to send the message m'. These changes can occur (with positive probability) while the strategies of the role two players remain fixed. Next, the role two players can switch (one by one) to a strategy that responds optimally to the first role's strategy. This strategy can be selected to have the property that role two players take action a_2 in response to an unsent message m'' (independent of the message that role two players themselves send). Next, change the strategy of every role one player so that they all send m'' and take action a_1 independent of the message that they hear from their opponent. Since u_1^* is the largest feasible payoff for a role one player, this replacement satisfies (NL) (indeed, it is an optimal response to the population strategy). At this point the population strategy profile is homogeneous and leads to the payoff u_1^* for role one players. To obtain the strategy θ^*, let role two players modify their strategies so that they all send the message m and take a_2 in response to every message; and let role one players modify their strategies (if $m'' \neq m$) so that they too send m.

Propositions 1 and 2 guarantee that in common-interest games there is a unique stable set, and it contains precisely those strategies that lead to efficiency. The logic behind the result is plain. If the population is outside of this set, it can drift to a profile in which there is an unused message, and agents can use this message to coordinate on a good outcome. When there are common interests, the population gets stuck once everyone obtains an efficient payoff. A direct consequence of these results is that payoffs converge globally in common-interest games. The payoff to an individual in role i must converge to the efficient payoff, u_i^*, with probability one.

Since all of the replacements used in the proofs of Propositions 1 and 2 satisfy (BR), existence and uniqueness of a stable set in common-interest games holds under (BR) (instead of (R)).

4. GENERAL GAMES

When the common-interest assumption fails, the set of outcomes that arise infinitely often is large. The next result, which follows from a simple modifica-

tion of the proof of Proposition 2, states that the stable set of any communication game must include the efficient frontier of the set of stage-game payoffs.

PROPOSITION 3: *For any game with pre-play communication with at least two messages there is a unique stable set. For any efficient payoff in the underlying game, the stable set contains a homogeneous strategy profile yielding that payoff.*

Rather than provide a proof of this result, we shall illustrate the result with the prisoner's dilemma game. Suppose that the population was playing a homogeneous profile in which every agent used the same message and cheated no matter what his opponent said. Through a sequence of replacements, the strategies of the role two individuals could change so that they all respond to an unsent message with the cooperative action. Finally, through a sequence of replacements, the strategies of the role one players can change so that these players are using the previously unsent message (but still cheating). As a result, the role one players receive their highest payoff (even though this payoff could not be achieved in a static equilibrium).[3] It is straightforward to show that the population can move with positive probability to a configuration in which players in role one receive their highest payoff from any initial specification of the population. This configuration is not a stable one: when role two players have an opportunity to adjust their strategy, they will cheat again. Hence the population moves from configurations that are preferred by one type of agent, to configurations that are inefficient, to configurations that are good for the other type of agent. Without providing more details about the transition probabilities, we cannot say how much time the process will spend at efficient outcomes. Our dynamic does not limit the set of possible predictions when the underlying game lacks common interests. The uncontrolled drift off the equilibrium path permitted by (R) leads to this result.

5. NONE-SIDED COMMUNICATION

In the literature that assumes language has a focal meaning, it is generally the case that if only one player is able to communicate, then that player is guaranteed to achieve his favorite outcome. This result holds in our framework.[4] Imagine that only players in the first role are able to send messages. The argument of Proposition 2 demonstrates that a population profile in which the agent who is able to speak receives his highest feasible payoff must be an element of every stable set. Hence there is only one stable set.

---

[3] If instead the role one players used the previously unsent message and then cooperated, a replacement that satisfies (R) but not (BR), the population would move to an outcome in which both players cooperate.

[4] The efficiency result did not hold in our earlier model, Kim and Sobel (1992).

We can also prove a version of the efficiency result of Section 3. If player one obtains his highest payoff only at a strict equilibrium, then the only stable population profiles must give rise to this equilibrium. As long as role one players receive their highest payoff, only other role one strategies that lead to the highest payoff can enter the population. When the equilibrium is strict, no role two strategy that lowers a role one player's payoff can enter the population. We summarize these results in Proposition 4.

PROPOSITION 4: *Assume that only players in role one may signal prior to playing the underlying game. Let w_1^* be the highest feasible payoff for a player in role one of the underlying game. There is a unique stable set of population strategies. w_1^* is a payoff to role one players associated with a homogeneous population profile contained in the stable set. If the only way to obtain the payoff w_1^* for role one players is in a strict equilibrium in the underlying game, then role one players must receive w_1^* at each element of the stable set.*

Proposition 4 holds if the assumption that the role-one player's equilibrium is strict is replaced by the weaker condition: $w_1^* = u_1(t_1^*, t_2^*)$ and t_2 an optimal response to t_1^* implies that $u_1(t_1^*, t_2) = w_1^*$. This condition holds in all common-interest games.

6. CONCLUSION

In this section we note limitations and extensions of our approach and briefly discuss related papers.

It is traditional in evolutionary models to view players as being drawn from a single population. Assuming that players are drawn from a single population leads to difficulties that we believe are artificial. Cheap talk is needed not only to provide a myopically attractive way to avoid inefficient strict equilibria, but must also be used to create asymmetries. Our basic results for common-interest games remain true in this setting nonetheless,[5] provided that there exists at least one unsent message for each individual in the population, and that we replace (S) with the more restrictive assumption that the probability that any individual has the opportunity to change his strategy at the end of a round is positive.[6] Assuming that there are at least two messages for each individual in the population guarantees that in any population strategy profile there is an unused message for each individual. When the population is finite, agents can use these extra messages to determine their roles in the game. In an infinite population, symmetric model, Schlag (1993) demonstrates that one cannot rely on the replicator dynamic to guarantee efficiency even in games where players' have identical payoffs.

[5]Details of the argument are available from the authors.
[6]Our previous results continue to hold under (BR) if (S) is strengthened in this way.

   Without placing more structure on the model, it is not possible to state how long it would take for the population to reach a stable set, but our proofs suggest that convergence could be slow. The particular set of replacements that we identify requires one side of the population to change its strategy while strategies of the individuals on the other side do not change. While other paths to the efficient outcome exist, we have not identified them, and the population could remain at an inefficient outcome for a long time. This intuition appears, in a different form, in Banerjee and Weibull (1993) and Schlag (1993). These papers demonstrate that inefficient outcomes can be dynamically stable with respect to the replicator dynamic.

   The richness assumption (R) permits us to conclude that stable sets contain efficient outcomes; it also leads to our conclusion that there are many recurrent outcomes in games that do not have common interests. We exploit the property, implied by (R) (or (BR)), that replacement strategies may have arbitrary responses to unsent messages. In particular, nothing prevents players from using a strongly dominated strategy of the underlying game in response to an unsent message. While we think that it is plausible to permit changes in off-the-equilibrium-path behavior, economic agents, even those with limited rationality, should be able to avoid certain responses. An earlier version of this paper (Kim and Sobel (1992)) studied the outcomes that satisfy the static stability condition developed by Swinkels (1992). This condition requires new strategies to be best responses to the population that arises after the entry. Kim and Sobel (1992) show that this notion of evolutionary stability forces efficiency in common-interest games provided that there are unsent messages. In contrast to the analysis of this paper, however, one could not guarantee the existence of unused messages without making another assumption in addition to common interests.[7]

   Robson (1990) considers the possibility of creating extra strategies in evolutionary games. These strategies play the same role communication does in our model. Robson demonstrates that adding a strategy forces cooperation in coordination games and destabilizes the inefficient outcome in the prisoner's dilemma.

   Matsui (1991) applies a variation of the Gilboa and Matsui (1991) idea of cyclically stable sets, which correspond roughly to the steady-states of best-response dynamics, to show that the only cyclically stable set in 2×2 common-interest games with pre-play communication contains only efficient equilibria. The efficiency result does not generalize to larger games because mixed strategies are permitted and there may be no guarantee that there exists an unused message that agents can use as a way to move to an efficient outcome.

   Sobel (1993) presents existence and efficiency results for common-interest games with pre-play communication using a static evolutionary stability concept. He obtains an efficiency result for general common-interest games assuming a finite population of players who use only pure strategies. The results in Section 3

_____

   [7]Bhaskar (1992) and Kim and Sobel (1992) demonstrate that the unsent messages may not exist when mixed strategies are permitted.

and 4 of this paper go beyond Sobel (1993) because they rely on an explicit dynamic process, permit a more general class of replacements, and do not require large message spaces. Sobel (1993) uses the 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 (Bhaskar (1992), Fudenberg and Maskin (1991), Kim and Sobel (1992), and Wärneryd (1991)) that apply static evolutionary stability equilibrium concepts to cheap-talk games.

Matsui and Rob (1991) and Nöldeke, Samuelson, and van Damme (1991) 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 have periodic opportunities to change their strategies; and that mistakes or mutations occur and cause the models to have a unique ergodic distribution, which they show converges to the set of efficient payoffs as the probability of mistakes goes to zero. Because mutations are permitted, the dynamics in these papers do not satisfy (NL). Our result that cheap talk forces efficiency holds more strongly in our model than in theirs in two senses. First the class of common-interest games strictly includes the coordination games that they study. Second our assumptions guarantee that once the population coordinates on an efficient outcome, it remains there.

Dept. of Economics, PHBA, University of Iowa, Iowa City, IA 52242, U.S.A.
                                    and
Dept. of Economics, University of California-San Diego, 9500 Gilman Dr., LaJolla, CA 92093, U.S.A.

REFERENCES

BANERJEE, A., AND J. WEIBULL (1993): "Evolutionary Selection with Discriminating Players," Institut für Höhere Studien, Wien.

BHASKAR, V. (1992): "Noisy Communication and the Evolution of Cooperation," Delhi University.

FARRELL, J. (1988): "Communication, Coordination, and Nash Equilibrium," *Economics Letters*, 27, 209–214.

FUDENBERG, D., AND E. MASKIN (1991): "Evolution and Communication in Games," preliminary notes, April.

GILBOA, I., AND A. MATSUI (1991): "Social Stability and Equilibrium," *Econometrica*, 59, 859–867.

KIM, Y.-G., AND J. SOBEL (1992): "An Evolutionary Approach to Pre-Play Communication," UCSD Discussion Paper.

MATSUI, A. (1991): "Cheap-Talk and Cooperation in a Society," *Journal of Economic Theory*, 54, 245–258.

MATSUI, A., AND R. ROB (1991): "The Role of Public Information and Pre-Play Communication in Evolutionary Games," University of Pennsylvania.

MYERSON, R. (1989): "Credible Negotiation Statements and Coherent Plans," *Journal of Economic Theory*, 48, 264–291.

NÖLDEKE, G., L. SAMUELSON, AND E. VAN DAMME (1991): "The Evolution of Communication," Universität Bonn.

RABIN, M. (1994): "A Model of Pre-Game Communication," *Journal of Economic Theory*, 63, 370–391.

ROBSON, A. (1990): "Efficiency in Evolutionary Games: Darwin, Nash and the Secret Handshake," *Journal of Theoretical Biology*, 144, 379–396.

SCHLAG, K. (1993): "Cheap Talk and Evolutionary Dynamics," Universität Bonn.

SOBEL, J. (1993): "Evolutionary Stability and Efficiency," *Economics Letters*, 42, 301–312.

SWINKELS, J. (1992): "Evolutionary Stability with Equilibrium Entrants," *Journal of Economic Theory*, 57, 306–332.

WÄRNERYD, K. (1991): "Evolutionary Stability in Unanimity Games with Cheap Talk," *Economics Letters*, 36, 375–378.