GAMES AND ECONOMIC BEHAVIOR 5, 547–575 (1993)

Evolutionary Stability in Games of Communication*

ANDREAS BLUME AND YONG-GWAN KIM

Department of Economics, University of Iowa, Iowa City, Iowa 52242

AND

JOEL SOBEL

Department of Economics, University of California, San Diego, La Jolla, California 92093

Received September 1, 1992

This paper identifies evolutionarily stable outcomes in games in which one player has private information and the other takes a payoff-relevant action. The informed player can communicate at little cost. Outcomes satisfying a set-valued evolutionary stability condition must exist and be efficient in common-interest games. When there is a small cost associated with using each message the outcome preferred by the informed player is stable. The paper introduces a nonequilibrium, set-valued stability notion of entry resistant sets. For games with partial common interest, the no-communication outcome is never an element of an entry resistant set. *Journal of Economic Literature* Classification Numbers: C72,D82.

1. INTRODUCTION

This paper identifies evolutionarily stable outcomes of communication games. We discuss simple Sender–Receiver games in which one player has private information and another player takes a payoff-relevant action.

* We have benefited from being able to read the related work of Canning (1992), Nöldeke and Samuelson (1992), and Wärneryd (1993) while preparing this manuscript. Comments by Antonio Cabrales, Vincent Crawford, Matthew Rabin, who deserves at least partial credit for the ideas in Section 7, Larry Samuelson, and two referees improved the paper. Blume and Kim thank the College of Business Administration at the University of Iowa, and Sobel thanks the NSF for financial support.

0899-8256/93 $5.00

The central issue of this paper is whether messages take on commonly understood meanings that permit the informed player to communicate effectively when it is in her interest to do so. Our motivation for using the evolutionary approach is that it enables us to identify situations in which talk can be used to eliminate inefficient outcomes without assuming that words have conventional meanings. We show that evolutionary pressures may force populations to interpret messages in systematic ways.

Our method of describing evolutionarily stable outcomes in communication games follows an earlier paper by two of us (Kim and Sobel, 1992), which characterizes the set of outcomes that satisfy a static evolutionary stability notion in two-player, normal-form games that have been augmented by one round of simultaneous communication. The general message of this work is that when preplay communication is possible, evolutionary pressures destabilize inefficient outcomes. The strongest existence and efficiency results come in games with common interests (that is, where the underlying game has a unique Pareto efficient point). Otherwise the existence of stable outcomes is in doubt. The topic has attracted the attention of several others. To date we are aware of work by Bhaskar (1992), Fudenberg and Maskin (1991), Matsui (1991), and Wärneryd (1991) that all arrive at roughly the same conclusions using roughly the same approach. These models provide some justification for the claim that communication leads to efficiency. On the other hand, when players have complete information about the game and the game has a unique efficient payoff, there is always an efficient Nash equilibrium; communication is not necessary to avoid inefficiency. Talking does serve the role of communicating strategic intent and destabilizing bad equilibria; however, players could have coordinated on efficient equilibria without talking, and once efficiency is attained there is no further need to talk prior to play. Here talk actually enlarges the set of possible outcomes because we assume that players have different information. Canning (1992), Nöldeke and Samuelson (1992),[1] and Wärneryd (1993) have already made contributions to this literature. Our paper uses a different solution concept and obtains more general results. We discuss some of the related literature in detail in the final section of this paper.

Section 2 describes our basic model. Section 3 presents two important preliminary results. Under our maintained assumption that the set of messages is large, the first result demonstrates that every stable set must contain a strategy that does not use some messages. The second result demonstrates that stable sets must also contain strategies that do not

[1] The published version of this paper (Nöldeke and Samuelson, 1993) does not contain the dynamic analysis of cheap-talk games which appeared in the working paper that we reference.

punish the use of unused messages. Taken together, these results identify how a set of strategies may become vulnerable to invasion. First, the population drifts to a state where some words are unused. Second, punishments (in the form of lower payoffs) that could be associated with using these messages disappear. Third, the informed players exploit the unused messages to move the population to a desirable outcome. Section 4 shows how the third step works in common-interest games. We show that outcomes which satisfy our stability condition must exist and be efficient. This result is the standard one for the literature. Section 5 relaxes the common-interest assumption. It discusses games in which there is an equilibrium where the informed player receives her highest feasible payoff. We show that if there is a small cost associated with using each message, and these costs differ from message to message, then the outcome preferred by the informed player is stable. We give an example to show that it need not be the only stable outcome. In Section 6 we compare our solution to cheap-talk refinements. We give a simple condition under which any outcome that fails the refinements introduced by Farrell (1993), Matthews *et al.* (1991), and Rabin (1990) cannot be evolutionarily stable. These results suggest that existence of stable outcomes is unlikely. In Section 7 we discuss limiting outcomes of games in which a stable set of strategies need not exist. We give conditions that rule out the no-communication outcome in a general set of games with partial common interest.

There has been an explosion of papers recently that use evolutionary arguments to rule out inefficient outcomes in games. We discuss the papers most similar to our work and related modeling issues in Section 8.

2. The Basic Framework

We confine our analysis to simple signaling games. The Sender has private information; the Receiver must take an action that is relevant to both players' payoffs. In this class of games, the ability to communicate influences the set of possible outcomes. Allowing the Sender to say something about her private information gives the Receiver a chance to condition his action on a (possibly noisy) signal of the state of the world. Without this signaling, the Receiver would be unable to take a state-contingent action.

Assume that prior to the game nature selects the Sender's type t from a finite set T according to a distribution $\pi(\cdot)$;[2] that the Sender's set of

[2] We could equivalently assume that there is a different group of players representing each type of Sender and that $\pi(t)$ is the proportion of type t Senders in the population.

pure strategies consists of rules that assign to each t a message m, which is a member of a finite set M; that the Receiver's set of pure strategies consists of rules that assign to each m an element a of a finite set A; and that the players have utility functions $u_i(t, m, a)$ for $i = 1$ (for the Sender) and $i = 2$ (for the Receiver). We further restrict the way in which signals enter payoff functions. We assume that $u_2(t, m, a) \equiv v_2(t, a)$ does not depend on m and that $u_1(t, m, a) \equiv v_1(t, a) - c_1(m)$, where $c_1(\cdot)$ represents the cost of signaling. We either assume that signaling is costless, $c_1(\cdot) \equiv 0$, or that costs are nominal in the sense that $c_1(\cdot)$ is small relative to other payoffs. We introduce our notion of games with nominal signaling costs in conjunction with a genericity assumption on payoffs.

We say that a cheap-talk game has *generic payoffs* if $v_1(t, a) = v_1(t, a')$ implies that $a = a'$ and the Receiver has a unique best response to any strategy sent with probability one by a nonempty subset of Sender types; that is, for each nonempty $T' \subset T$, $\arg\max_{a \in A} \sum_{t \in T'} v_2(t, a)\pi(t)$ is a single action. We say that a cheap-talk game has *nominal signaling costs* if it has generic payoffs and the costs of messages can be ordered

$$0 \leq c_1(m_1) < c_1(m_2) <, \ldots, < c_1(m_i) <, \ldots, < c_1(m_k) < \delta,$$

where $\delta = \min \{|v_1(t, a) - v_1(t, a')|/2: t \in T, a \neq a'\}$. The genericity assumption guarantees that $\delta > 0$. Assuming that all of the messages cost less than the smallest difference in payoffs in the basic game is in keeping with the idea that talk is cheap. Canning (1992) uses a related assumption in his work.

A mixed strategy is a pair $\sigma = (\sigma_1, \sigma_2)$, where $\sigma_1(m, t)$ is the probability that a Sender of type t sends the message m, and $\sigma_2(a, m)$ is the probability that the Receiver takes the action a in response to the message m. A strategy σ gives rise to a payoff $U(\sigma) = (U_1(\sigma), U_2(\sigma))$, where $U_i(\sigma) = \sum_{t \in T} \sum_{m \in M} \sum_{a \in A} u_i(t, m, a)\sigma_1(m, t)\sigma_2(a, m)\pi(t)$ for $i = 1$ and 2. σ is a Nash equilibrium if the strategies respond optimally to one another:

$$\text{if } \sigma_1(m, t) > 0, \quad \text{then } m \text{ solves } \max_{m' \in M} \sum_{a \in A} u_1(t, m', a)\sigma_2(a, m') \quad (1)$$

and

$$\text{if } \sigma_2(a, m) > 0, \quad \text{then } a \text{ solves } \max_{a' \in A} \sum_{t \in T} u_2(t, m, a')\sigma_1(m, t)\pi(t). \quad (2)$$

We assume throughout that there are enough elements in M to enable the informed player to use an unsent message to avoid a bad outcome. We assume that

$$\#(M) > 2^{[\#(A) + \#(T)]} + \#(T), \quad (3)$$

where $\#(X)$ is the cardinality of the set X. We will show in Section 3 that if condition (3) holds, then any stable set must contain a strategy that does not use at least $\#(T)$ messages. The interesting applications of cheap-talk models involve large message spaces; assuming (3) does not rule out anything of importance.

As in Kim and Sobel (1992), we use as our basic stability concept Swinkels's (1992a) equilibrium evolutionarily stable (EES) sets. An EES set is a nonempty, closed set of Nash equilibria that is stable against a certain class of invasions. The allowable invasions must not only be optimal responses to the population strategy, but also to the population strategy that results after the entry of a small group of invaders. This definition differs from the standard definition of ESS (Maynard Smith, 1982, or Maynard Smith and Price, 1973) because it directly applies to asymmetric games, because it looks for stable sets rather than stable strategies, and because it places restrictions on the set of possible entrants. Swinkels (1992a,b) shows that every EES set contains a proper equilibrium and satisfies the never-a-weak-best-response property of Kohlberg and Mertens (1986). Some modification of ESS is needed to avoid the trivial nonexistence problems it has in games with unreached information sets and redundant strategies. Other ideas have been presented by Fudenberg and Maskin (1990, 1991), Gilboa and Matsui (1991), Hofbauer and Sigmund (1988), Selten (1983), and Thomas (1985a,b). The crucial modeling decisions are whether or not the stability condition should be set valued, and whether or not there should be some restrictions on the kind of strategies that might enter the population. We choose a set-valued solution concept because it enables the population strategies to drift off the equilibrium path and thereby makes it more difficult for a population to remain at an inefficient outcome when a Pareto-dominating equilibrium exists. Efficiency results for single-valued solution concepts hold only under much stronger assumptions than ours (see Wärneryd, 1993). We choose to restrict entry because adding a dominated strategy to a game could otherwise change the set of predictions. Section 8 contains a more detailed discussion of alternative modeling approaches.

We state Swinkels's definition for a general two-player game with strategy set $S = S_1 \times S_2$ and payoff functions $u = (u_1, u_2)$, which we represent by (S, u). Let $\mathcal{N}(S, u)$ be the set of Nash equilibria of (S, u); let $\Delta(S_i)$ be the set of mixed strategies of player i; let $C(s)$ be the carrier of s (the set of pure strategies given positive probability by s); and let $\mathrm{BR}_i(\cdot)$ be the best response correspondence of player i for $i = 1$ and 2. For $\sigma = (\sigma_1, \sigma_2) \in \Delta(S_1) \times \Delta(S_2)$, we set $\mathrm{BR}(\sigma) \equiv (\mathrm{BR}_1(\sigma_2), \mathrm{BR}_2(\sigma_1))$.

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 conditions,

There exists $\varepsilon' \in (0, 1)$ such that for all $\varepsilon \in (0, \varepsilon')$ and for all $\sigma \in \Theta$, if $C(\sigma') \subset \mathrm{BR}((1 - \varepsilon)\sigma + \varepsilon\sigma')$, then $(1 - \varepsilon)\sigma + \varepsilon\sigma' \in \Theta$. (4)

$$\Theta \text{ is closed and nonempty.} \tag{5}$$

$$\Theta \subset \mathcal{N}(S, u). \tag{6}$$

Conditions (5) and (6) in the definition are familiar. They require that Θ be a closed set of Nash equilibria. Condition (4) 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. The definition permits individuals in the population to play mixed strategies. Unlike other variations on the static ESS concept adapted to asymmetric games, a mixed strategy, taken as a singleton, may be an EES set. All of the EES sets that we describe in this paper include a pure-strategy equilibrium.

The important difference between the ESS and EES conditions is that admissible invasions in the EES framework are calculated to respond optimally to the population mixture that they induce rather than being simply random events. Even though EES is substantially weaker than ESS, there is no guarantee that EES sets exist. To obtain a general existence result, condition (6) must be abandoned. We do that in Section 7.

3. Preliminary Results

In this section we present some preliminary lemmas that indicate the power of evolutionary stability concepts in communication games.

First we show that any EES set must contain a strategy that does not use some messages.

Lemma 1. *Any EES set in a cheap-talk game contains a strategy in which the Sender assigns probability zero to at least #(T) messages.*

Proof. Let $\sigma \in \Theta$. Denote the messages $m_1, m_2, \ldots, m_k$, and call message m_i redundant if either $\sigma_1(m_i, t) = 0$ for all t, or if there exists $j < i$ such that

$$\{t: \sigma_1(m_i, t) > 0\} = \{t: \sigma_1(m_j, t) > 0\} \tag{7}$$

and

$$\{a: \sigma_2(a, m_i) > 0\} = \{a: \sigma_2(a, m_j) > 0\}. \tag{8}$$

In words, (7) and (8) state that there are a pair of messages that induce

precisely the same set of responses that are sent with positive probability by precisely the same set of types. Call the pair of messages m_i and m_j equivalent if (7) and (8) hold.

By assumption (3) on the cardinality of M, it follows that there are at least $\#(T)$ redundant messages. It suffices to show that there is $\sigma' \in \Theta$ such that $\sigma_1'(m, t) = 0$ for all t and for all redundant messages. We can do this by having the Sender remove all weight from any redundant message and instead use a message equivalent to it.

Let $E(m_i)$ denote the equivalence class containing the message m_i; $E(m_i)$ is the set of all messages m_j for which (7) and (8) hold. Let $I(m)$ be the minimum index message in the class. Let $\sigma_2'(\cdot) \equiv \sigma_2(\cdot)$ and let $\sigma_1'(m', t) = 0$ if $m' \neq I(m')$, and $\sigma_1'(m', t) = \sum_{m \in E(m')} \sigma_1(m, t)$ otherwise. Since $\sigma_1(\cdot)$ was an optimal response to $\sigma_2(\cdot)$, any type t for which $\sigma_1(m, t) > 0$ is indifferent between all messages in $E(m)$; hence $\sigma_1'(\cdot)$ is a best response to $\sigma_2'(\cdot) \equiv \sigma_2(\cdot)$. Since for all $m \in E(m')$, the same pure strategies are in the support of $\sigma_2'(\cdot, m)$, and each action in the support of $\sigma_2'(\cdot, m)$ is an optimal response to $\sigma_1'(m, \cdot)$, $\sigma_2'(\cdot)$ is a best response to $\sigma_1'(\cdot)$. Therefore σ' is a Nash equilibrium and $C(\sigma') \subset \mathrm{BR}(\sigma)$. The lemma now follows from Proposition 1 of Kim and Sobel (1992), which shows that if Θ is an EES set for (S, u), $\sigma \in \Theta$, $\sigma' \in \mathcal{N}(S, u)$, and $C(\sigma') \subset \mathrm{BR}(\sigma)$, then $\sigma' \in \Theta$.

It is clear that without enough messages, it will not be possible to apply evolutionary arguments to reach efficiency. Earlier evolutionary approaches to communication in games have needed to confront the issue. One approach, taken by Canning (1992) and Nöldeke and Samuelson (1992), is to restrict attention to a finite population in which each individual plays a pure strategy and to assume that the message space is large relative to the population size and the number of types of Sender. Under these assumptions every equilibrium must have unused messages. We have demonstrated the existence of unused messages from an assumption about the size of the language without explicitly ruling out randomization.

LEMMA 2. *Let Θ be an EES set in a cheap-talk game with costless signaling. Let $\sigma \in \Theta$ and let $\overline{m} \in M$ be used with probability zero. For any $m' \neq \overline{m}$, the strategy σ' defined by $\sigma_1'(\cdot) \equiv \sigma_1(\cdot)$, $\sigma_2'(\cdot, m) \equiv \sigma_2(\cdot, m)$ whenever $m \neq \overline{m}$, and $\sigma_2'(a, \overline{m}) = \sigma_2(a, m')$ is an element of Θ.*

Lemma 2 is the drift lemma. It states that if there is a message that is not used by some strategy in a stable set, then the response to that message can drift freely to any other response that supports the equilibrium. It follows from the definition of EES sets.

Lemma 2 deserves careful consideration. If the general population repels invaders who attempt to use a new message, then evolutionary pres-

sures need not lead to efficiency. Our solution concept allows the population to drift off the equilibrium path; invaders who differ from the general population only in the way that they respond to unused messages can enter and ultimately change the strategy used by other members of the population. We implicitly assume that the population will spend enough time at any equilibrium contained in a candidate stable set that a viable invasion, if it exists, will be able to take hold. Hence while a stable outcome may be an appropriate long-run prediction, for certain initial conditions it may take a very long time to arrive at a stable set.

If Θ is an EES set for a game with costless signaling, then it follows from Lemmas 1 and 2 that there exists an element of Θ in which there are unused messages that an invading strategy can use without lowering its payoffs. We use this property several times to demonstrate the efficiency of EES sets. When signaling costs are nominal, Lemma 2 need not be true (if $c_1(\overline{m}) < c_1(m')$, then σ' is not an equilibrium strategy since all types will strictly prefer the cheaper message).

Since both Lemma 1 and Lemma 2 describe conditions that EES sets necessarily satisfy, the Lemmas continue to be true for evolutionary equilibrium concepts that require stability against a larger class of invasions (for example, evolutionarily stable strategies or the ES sets described in Section 8).

4. Common-Interest Games

In this section we study games in which the players have similar preferences in the underlying game. We say that a game has *common interests* if the set of feasible expected payoffs has a unique Pareto-efficient point. That is, there exists a feasible payoff pair $u^* = (u_1^*, u_2^*)$ such that for any strategy σ, either $U(\sigma) = \sigma^*$ or $U_i(\sigma) < u_i^*$ for $i = 1$ and 2.

These games have received attention in other studies of communication (Blume and Sobel, 1991; Canning, 1992; Kim and Sobel, 1992; Matsui, 1991; Nöldeke and Samuelson, 1992; Rabin, 1990; and Wärneryd, 1993).[3] Common-interest games are a natural place to look for effective communication. In this class of games players should coordinate on an efficient equilibrium.

This section contains two basic results: Proposition 1 states that any stable payoff of a common-interest game with costless signaling must be efficient. Proposition 2 states that there exists an EES set for common-

[3] Canning (1992), Nöldeke and Samuelson (1992), and Wärneryd (1993) limit attention to a subset of common-interest games, while Blume and Sobel's (1991) definition is slightly more general than the one we use in this paper.

	A	B	C
t_1	2,3	0,0	2,2
t_2	0,0	2,3	2,2

FIGURE 1

interest games with costless signaling. We also include an example that demonstrates the importance of our common-interest assumption for proving existence. We conclude the section with the observation that our results do not change when messages have nominal costs.

PROPOSITION 1. *Let G be a costless signaling game of common interest. If Θ is an EES set of G, then for all $\sigma \in \Theta$, $U(\sigma) = u^*$.*

The proof of Proposition 1 has three steps. The first step, which we proved in Lemma 1, shows that there is always an element of an EES set that contains unused messages. The second step, which follows from repeated applications of Lemma 2, shows that if there exists an element of an EES set that contains an unused message, then there is also an element of the set that does not punish the use of the unsent message, in the sense that a type could send it without lowering its payoff. The third step demonstrates that a strategy which uses the unsent messages to coordinate on the efficient outcome can invade the population. We provide the details in the Appendix.

Proposition 2 proves that EES sets exist in common-interest games. We omit the proof since Kim and Sobel (1992) use the same argument to establish existence of EES sets in complete-information common-interest games.

PROPOSITION 2. *If G is a costless signaling game of common interest, then $\Theta = \{\sigma: U(\sigma) = u^*\}$ is an EES set of G.*

The intuition for Proposition 2 is that, by the common-interests assumption, any strategy profile that responds optimally to a strategy in Θ must also be an element of Θ. Hence Θ is an EES set.

Our definition of common interests requires that if one player obtains its highest payoff, then so must the other. We need this assumption in the existence theorem. Consider Example 1 in Fig. 1. In all of our examples the entry in row i and column j indicates the payoffs to the Sender of type i and to the Receiver if j is the Receiver's action; all types of the Sender are equally likely. In this game the Sender and the Receiver do not have common interests by our definition because the Sender is indifferent between the separating and the pooling equilibrium. No EES set exists for

the example. To see this assume, in order to obtain a contradiction, that there exists an EES set. We can show that the set must contain the equilibrium in which the Sender always sends m_3, and the Receiver responds to m_1 with a mixture of A and C, to m_2 with B, and to m_3 with C. The equilibrium in which type t_1 Sender signals m_1, type t_2 Sender signals m_2, and the Receiver responds to m_1 with A, to m_2 with B, and to m_3 with C can now enter the population. When it does, the population's response to m_1 is not optimal. Hence the population strategy is not an equilibrium, so we did not start with an EES set.

Talk may be cheap but it is rarely literally free. We close this section by observing that even if there is a small cost associated with messages, the basic results of this section continue to hold. Canning (1992) presents versions of Propositions 3 and 4 below in a dynamic model.

Variations of Propositions 1 and 2 also apply to games with nominal signaling costs. Proposition 3 is the analog to Proposition 1. We omit the proof, which differs slightly from the proof of Proposition 1 because Lemma 2 does not extend to games with nominal signaling costs. We must guarantee that an equilibrium in which the Sender always obtains her most preferred action can enter the population. If a Sender type uses a message that fails to induce her utility maximizing action with probability greater than one-half, then there exists a message that is not used and is such that she is indifferent between her equilibrium message and the new message even when nominal signaling costs are taken into account. This construction is possible because the new message can induce the Sender's favorite action with a high enough probability to compensate for the added signaling charge. Consider a message m that induces the Receiver to take the utility maximizing action with probability greater than one-half for a nonempty set of Sender types. There can only be one action, call it a, induced with probability greater than one-half by the message m. By the common-interest assumption, a will be a best response if only the Sender types for which a is the utility maximizing action use m. Hence, at any equilibrium in which the Sender is not obtaining her maximum utility (less signaling costs), an invading strategy in which each type of Sender uses a message from the equilibrium if it induces her favorite action with probability greater than one-half and a previously unused message otherwise, and in which the Receiver responds optimally to the Sender's invading strategy, can enter the population.

PROPOSITION 3. *Let G be a common-interest cheap-talk game with nominal signaling costs. If Θ is an EES set of G, then for all $\sigma \in \Theta$, $\sigma_1(t, m_i) = 0$ for $i > \#(T)$, $U_1(\sigma) \geq u_1^* - \delta$, and $U_2(\sigma) = u_2^*$.*

Proposition 3 modifies the efficiency result of Proposition 1. It asserts that only the cheapest messages will be used; this is true because otherwise

an invading strategy using a cheaper message could enter. Proposition 1 places no restriction on the number of messages used in an element of an EES set. On the other hand, Proposition 3 does not guarantee that the Sender uses messages efficiently: It is possible that she uses the cheapest message to convey rare information and uses more expensive messages with higher probability. The Sender does not obtain her maximum payoff u_1^* because she must pay the nominal cost of signaling.

Proposition 2 also has a counterpart with nominal signaling costs. To state the proposition we need another definition. Given any strategy profile σ, denote by $\#(\sigma)$ the cardinality of the set of actions induced with positive probability under σ. That is, $\#(\sigma)$ is the cardinality of $\{a \in A: \sum_{m \in M} \sigma_2(a, m)\sigma_1(m, t)\pi(t) > 0 \text{ for some } t \in T\}$.

PROPOSITION 4. *If G is a common-interest cheap-talk game with nominal signaling costs, then every equilibrium σ in which $U_2(\sigma) = u_2^*$ and in which the Sender uses only the cheapest $\#(\sigma)$ messages is an element of an EES set of G.*

The proposition asserts that any equilibrium which is efficient from the standpoint of the Receiver and does not needlessly use costly messages (in the sense that the Sender uses only one message per distinct action induced in equilibrium) is an element of an EES set. In common-interest games it is natural to expect that the equilibria which the Receiver prefers are fully separating so that $\#(\sigma) = \#(T)$. Our statement of the proposition includes degenerate cases in which a completely informed Receiver has the same optimal response to two different types of Sender, so that fully separating equilibria do not exist. In contrast to Proposition 2, which asserts that the set of all strategies giving rise to the efficient equilibrium are elements of a single EES set, there are many different EES sets when signaling costs are nominal. These sets differ depending on which type uses which message. All that we can assert is that the Sender will use only the cheapest messages. In each element of an EES set the Sender has a unique optimal response. Since messages are costly and only the cheapest messages are used, each type strictly prefers her equilibrium message to any unused message. Each of the EES sets described in Proposition 4 contains all of the equilibria that support a given separating outcome.

5. CHOOSING THE SENDER'S FAVORITE EQUILIBRIUM

In many models of communication there is a presumption that if only one player is able to speak, then that player will be able to select its favorite outcome. Most of the solutions defined for this class of games

	A	B	C
t_1	1,3	0,0	2,2
t_2	0,0	1,3	2,2

FIGURE 2

give the Sender the ability to pick her favorite equilibrium under certain conditions (for example, Matthews *et al.*, 1991; Myerson, 1989; Rabin, 1990; and Zapater, 1991). Since only the Sender can speak in our model, we investigate the sense in which only her favorite outcomes are evolutionarily stable. First we need to understand what it means to be the Sender's favorite outcome. Since the type of the Sender determines her preferences, there may be conflicts of interest between the different types. We will only consider cases in which the interests of the Sender are completely clear. We say that a cheap-talk game has a *favorite equilibrium* for the Sender if there exists a Nash equilibrium of the costless signaling version of the game in which each type of Sender receives her highest payoff. When costs are nominal, Sender types will disagree about who uses the cheapest messages.

We begin our discussion with a negative result. Consider the example in Fig. 2. In this game there is a pooling equilibrium and a separating equilibrium. The Sender types would both prefer to pool, but the Receiver would prefer the Sender to use different messages for different types. When all signaling is free, there does not exist an EES set. To see that the pooling equilibrium outcome is not stable, first note that any EES set that contains the pooling outcome contains a strategy in which the Receiver is indifferent between actions A and C given one message, and indifferent between actions B and C given another message. Hence the separating equilibrium (with the type t_1 Sender using the first message and the type t_2 Sender using the second message) can invade. On the other hand, the pooling strategy can invade a population that is playing the separating equilibrium by using a different message provided that the invader does not get punished by the original population for doing so. It can be checked that any EES set containing the separating equilibrium will contain a strategy that does not punish invaders.[4]

The example suggests that there is a severe problem in establishing the

[4] Blume (1992) applies his concept of perturbed message persistence to the game in Fig. 2. His discussion parallels ours. Both the separating and the pooling outcomes are perturbed message persistent. However, under an effective language requirement, which plays the role of our assumption of nominal signaling costs, only the pooling outcome survives

existence of EES sets in games without common interests. The problem is not as bad as it seems, however. The strategy that invades the pooling equilibrium requires that the population mix its messages in a precise way. This type of mixture is possible because all messages cost the same amount (they are all free). Consider the effect of adding a cost of signaling that is independent of type and action, but varies from message to message. The modification guarantees that the pooling outcome is an element of an EES set. Furthermore, in the EES set all Senders in the population choose to use only the cheapest messages. The next result shows that this is a general property.

PROPOSITION 5. *Let G be a cheap-talk game with nominal signaling costs. If the Sender has a favorite equilibrium in G, then there exists an EES set in which the Sender obtains her highest payoff (less signaling costs) for each element in the set.*

Proof. Since the game has generic payoffs, any equilibrium in which the Sender obtains her highest payoff is a pure-strategy equilibrium along the equilibrium path. The Receiver cannot randomize on the equilibrium path and give the Sender her highest payoff because $v_1(t, a) \neq v_1(t, a')$ whenever $a \neq a'$. Given that the Receiver is playing pure strategies, no type of Sender will be indifferent between two messages. Let σ be an equilibrium that gives rise to the Sender's favorite payoff in which the Sender uses only the cheapest messages with positive probability ($\sigma_1(m_i, t) = 0$ for all t implies that, for all $j > i$, $\sigma_1(m_j, t) = 0$ for all t). Let Θ be the set of equilibria that give rise to the same outcome (elements in Θ differ only in the way that they respond to unused messages). Θ is an EES set because, by construction, σ_1 is the unique best response to any strategy for the Receiver in Θ.

While Proposition 5 asserts that the Sender's favorite outcome is part of a stable set, it does not guarantee that it is the only stable outcome. The next example demonstrates that such an efficiency result is not available.

In Example 3 there is a partially separating equilibrium in which actions A and B are taken. The Sender prefers to pool, but the partially separating equilibrium is an element of an EES set. Consider the set $\Theta = \{\sigma \in \mathcal{N}: \sigma_1(m_1, t_1) = \sigma_1(m_2, t_2) = \sigma_1(m_2, t_3) = 1, \sigma_1(\cdot) = 0, \text{otherwise}\}$. We claim that Θ is an EES set. We will sketch an argument that demonstrates that no strategy can invade Θ. First, observe that if σ_1' is an optimal response to any σ_2 such that $\sigma = (\sigma_1, \sigma_2) \in \Theta$, then $\sigma_1'(t_3) = 1$. Type t_3 can get a payoff of four (less signaling costs) if he uses m_2. Plainly t_3 would never use m_1, which would lead to a negative payoff when signaling costs are taken into account. Moreover, the Receiver's response to any of the other messages must support the equilibrium. It is straightforward to check that

	A	B	C
t_1	4,10	0,0	10,4
t_2	1,0	4,8	11,4
t_3	0,0	4,2	10,4

FIGURE 3

any response to m_i for $i > 2$ that would leave t_3 indifferent between m_2 and m_i would cause t_2 to strictly prefer m_i to m_2 and destroy the equilibrium. It follows that for any potential invading strategy, only types t_1 and t_2 can use messages m_i for $i > 2$. Consequently, if σ' can invade Θ, then $\sigma_2'(C, m_i) = 0$ for all $i > 2$. That is, since t_3 never sends a new message, invaders that use the pooling action in response to a new message cannot enter. It follows that Θ satisfies condition (4) in the definition of EES. Hence the nonpooling equilibrium outcome is evolutionarily stable in the game of Fig. 3.

The outcome that the Sender dislikes can be stable in the example because it is possible to support the outcome with strategies that always impose a cost on an invasion that uses strategy C.

As a referee has pointed out to us, restricting entry to strategies that respond optimally to the postentry environment plays an essential role in guaranteeing that the semipooling equilibrium is an element of an EES set in the example. It is this restriction that prevents the drift of the population to strategies outside an equilibrium component. If entry were unlimited, then strategies could drift arbitrarily at unreached information sets; the drift would destabilize the semipooling equilibrium in the example. This type of drift is permitted in the dynamic model of Nöldeke and Samuelson (1992) and in the ES sets of Thomas (1985a,b).

It is possible to show that if there are only two Sender types, three actions for the Receiver, and the Sender has a favorite equilibrium, then the Sender must obtain her highest payoff in any EES set. This result corresponds to a finding in Nöldeke and Samuelson (1992). The result appears to be quite special in our model, as it does not hold when the Sender has more than two types or when the Receiver has more than three actions.

6. RELATIONSHIP TO REFINEMENTS

Several authors have presented equilibrium refinements in attempts to rule out implausible outcomes in costless signaling games. In this section we make a partial comparison between their results and ours.

We present a result based on the lemmas in Section 2. This result allows us to show when an equilibrium that fails the refinements of Farrell (1993), Matthews *et al.* (1991), and Rabin (1990) also fails to be an element of an EES set.

For $t \in T$, let $w(t)$ be a collection of reference payoffs. Call a nonempty subset J of T a *self-signaling family* relative to w if there exists a partition of J into sets J_i, $i = 1, \ldots, j$ and actions a_i^* such that

$$a_i^* \text{ solves } \max_{a \in A} \sum_{t \in J_i} v_2(t, a)\pi(t) \quad \text{for } i = 1, \ldots j,$$

$$v_1(t, a_l^*) > \max \{w(t), v_1(t, a_i^*)\} \quad \text{for } t \in J_l \text{ and } i \neq l, \text{ and}$$

$$v_1(t, a_i^*) < w(t) \quad \text{for all } i \text{ if } t \notin J.$$

Call J a *self-signaling set* if $j = 1$. Informally, a subset J is self signaling if there is an optimal response to the statement "my type is in J_i" that precisely those types in J_i prefer relative to reference payoffs. Farrell (1993) and Matthews *et al.* (1991) derive the reference payoffs from the equilibrium that they are testing for viability. Farrell calls an equilibrium *neologism proof* if there is no self-signaling set relative to the Sender's equilibrium payoffs.

Matthews *et al.* (1991) present two variations on Farrell's ideas, which they call announcement-proof and strongly announcement-proof equilibria. Their ideas differ from Farrell's in three ways. First, they do not assume that the deviating Sender types (the set J) can convince the Receiver to take any optimal response to J. If there exist multiple optimal responses, then they apply the most stringent credibility conditions: types in J expect their least-preferred response while types outside of J expect their most-preferred response to a neologism. Second, many new messages can be sent at the same time. That is, the self-signaling family contains more than one member. Third (for announcement-proof equilibria), an additional credibility condition is imposed to guarantee that if there are multiple statements that might be believed, their interpretation creates no conflicts in the sense that no matter what the Receiver believes, it is in the interest of the putative deviators to leave the equilibrium. See Matthews *et al.* (1991, p. 261) for a detailed discussion of this condition. Matthews *et al.* (1991) permit randomization in their announcements. We choose to limit attention to pure strategies for clarity. We believe that our results extend to their more general setting.

We will use the next proposition to compare these solution concepts to equilibrium evolutionary stability.

PROPOSITION 6. *Let σ be a pure-strategy Nash equilibrium of a cheap-talk game with costless signaling and generic payoffs. If $J = \bigcup_{i=1}^{j} J_i$ is a self-signaling family relative to $U_1(\sigma)$ and all types in J_i send the same*

message under σ *(if* t *and* $t' \in J_i$, *then* $\sigma_1(m, t) = \sigma_1(m, t')$*), then* σ *is not an element of an EES set.*

We present a proof of Proposition 6 in the Appendix. Assume that σ satisfies the conditions of the proposition and is an element of an EES set. Lemmas 1 and 2 imply that there exists another element of the EES set containing σ in which there exists an unused message m_i for each element of the self-signaling family, and that Sender types in J_i are indifferent between using their equilibrium message and m_i. It follows that a strategy in which Sender types in J_i send a common message and in which the Receiver responds optimally to this message can enter the population. When there are nominal signaling costs it may not be possible to simultaneously compensate several types of the Sender with the amount needed to leave them indifferent between their original strategy and sending the unused message. Hence our proof of Proposition 6 does not hold for these games.

Proposition 6 gives a condition under which an equilibrium that fails to be neologism proof or (strongly) announcement proof must also fail to be evolutionarily stable. Evolutionary pressures disrupt an equilibrium that fails a refinement provided that the types pooling together to make a credible statement forbidden by the refinement pool together in the equilibrium. Of course, this condition is satisfied automatically if the candidate equilibrium is pooling. If the members of an element of the self-signaling family use different messages in the candidate equilibrium, then an outcome that is not neologism proof or (strongly) announcement proof may be an element of an EES set. One can see this by examining the game in Fig. 3 of Section 5. The semipooling equilibrium in that game is an element of an EES set even when messages are costless. However, relative to that equilibrium the set of all types is self-signaling; hence the equilibrium fails to be neologism proof and announcement proof. The reason for the difference is that in order for evolutionary forces to destabilize an outcome it must be profitable for deviants to use a previously unused message without being punished. Since the preinvasion population must respond to the new message with an action that supports the equilibrium it may not be possible to find an alternative message that the types of a self-signaling set are all willing to use when they are not using the same message.

We noted three differences between announcement-proof and neologism-proof equilibria. None of these differences plays an important role in our arguments. In particular the additional credibility restriction is not needed in order to apply evolutionary arguments to destabilize an outcome. Nöldeke and Samuelson (1992) make a related observation in their discussion of forward induction and evolutionary stability.

Proposition 6 gives only a sufficient condition for instability. We can show that if there are only two types of Sender, then any equilibrium that fails to be neologism proof, announcement proof, or a credible message equilibrium (Rabin, 1990) cannot be an element of an EES set.[5]

This section presents two ways in which refinements may differ from evolutionary arguments. First, the existence of a self-signaling family necessarily destabilizes an outcome only if types pooling in the invasion were pooled before the invasion. The reason for the differences is that according to the refinements, players immediately interpret a new message (that is, a message used with probability zero in the putative equilibrium) according to its focal meaning. If a refinement permits a deviation from an equilibrium, then the Receiver always selects an optimal response to the information contained in a message sent by deviating Sender types, even when this response is very different from the action that supported the putative equilibrium. Hence one tests the credibility of a deviation by comparing payoffs in the putative equilibrium to payoffs when an unused message is interpreted according to its focal meaning. To move away from an outcome in our framework the response to a new message must first drift away from the population's preinvasion action. Entry of a new strategy occurs only if it is in the short-term best interest of an invader to enter. Unless we make an assumption on the population strategies prior to the invasion, there is no guarantee that it will be in the short-term interest of all Sender types in an element of a self-signaling family to send a common message and thereby allow evolutionary pressures to establish the meaning of this message.

The second difference is that many of the nonevolutionary ideas (Matthews *et al.*, 1991; Rabin, 1990; and Zapater, 1991) require that there be only one possible way to interpret a message. Again we can trace the difference to the unsophisticated behavior that is implicit in the evolutionary solution concept. An invasion takes hold slowly, myopically, and without introspection. When a new strategy designed as in Proposition 6 enters the population, it is able to grow. As it grows, an unambiguous interpretation of the previously unused messages develops without reference to other interpretations that may have been possible. Players who had more ability to reason about the game could easily fail to respond correctly to messages that had several conceivable credible meanings. The approach of Matthews, *et al.* (1991), Rabin (1990), and Zapater (1991) could be appropriate in dynamic settings with thoughtful agents.

[5] Nöldeke and Samuelson (1992) show that for games with two types and three actions, limiting outcomes of their evolutionary dynamic must be announcement proof whenever announcement-proof equilibria exist.

7. Partial Common Interest

The results of the previous sections strongly suggest that nonexistence of EES sets is widespread in games with communication. In this section we discuss a weaker stability notion and identify a class of games in which partial communication necessarily occurs.

Our approach is simply to eliminate condition (6) in the definition of EES.

Definition. A set $\Theta \subset \Delta(S_1) \times \Delta(S_2)$ is *entry resistant* (ER) if it is minimal with respect to (4) and (5).

It is clear from the definition that any EES set is an ER set. Two other things should be noted about the definition. First, ER sets may contain nonequilibrium strategies (indeed, they may contain only nonequilibrium strategies). Second, a basic question is whether there exist explicit dynamics that have ER sets as their limit points. We do not have a definite answer to that question as we provide no dynamics in this paper.

ER sets are closely related to the cyclically stable sets proposed by Gilboa and Matsui (1991) and Matsui (1992).[6] One can establish the existence of ER sets for all games using a Zorn's Lemma argument. (Gilboa and Matsui, 1991, and Matsui, 1992, give details of closely related arguments.[7])

Our purpose for introducing ER sets is to provide conditions under which evolutionary pressures guarantee that some communication takes place even though an EES set may not exist. Proposition 7, which we state below, provides conditions under which no pooling equilibrium (an equilibrium that gives rise to the same outcome as a completely uninformative equilibrium $\rho = (\rho_1, \rho_2)$ in which $\rho_1(m, t)$ does not depend on t) can be an element of an ER set. At the end of the section we discuss and provide conditions which guarantee that an arbitrary equilibrium strategy belongs to no ER set. This result allows us to conclude that in common-interest games the efficient equilibrium is the only equilibrium that is an element of an ER set.

In order to state our result, we must identify games in which the Sender has a compelling interest to share some, but not necessarily all, of her private information with the Receiver. For any nonempty subset of types L we write $\mathrm{BR}_2(L) \equiv \{\arg\max \sum_{a \in A} \sum_{t \in L} v_2(t, a)\mu(t)\alpha(a)$: α and μ are probability distributions supported on A and L, respectively$\}$ and $\mathrm{BR}_2(L$;

[6] The game that Matsui (1992, p. 358) uses to distinguish the definition of cyclically stable sets given in Gilboa and Matsui (1991) from that of Matsui (1992) also demonstrates that an ER set need not be cyclically stable in the sense of Gilboa and Matsui. We could not prove that ER sets were identical with Matsui's (1992) cyclically stable sets.

[7] Kalai and Samet (1984) employ a similar argument to establish existence of persistent retracts.

$\pi) \equiv \{\arg\max \sum_{a\in A} \sum_{t\in L} v_2(t, a)\pi(t)\alpha(a)$: α is a probability distribution supported on $A\}$; $\mathrm{BR}_2(L)$ contains all mixed strategies that an optimizing Receiver would consider assuming that the Sender's type is an element of L, while $\mathrm{BR}_2(L; \pi)$ requires the Receiver to derive the relative probabilities of the types in L from the prior π. Also let $\underline{v}_1(t; L) \equiv \min \{v_1(t, a_i)$: $a_i \in \mathrm{BR}_2(L)\}$ denote the lowest payoff that the type t Sender would obtain if the Receiver believed that her type was in L, and let $v_{1\mathrm{p}}(t)$ denote the maximum payoff that a type t Sender can obtain in a (completely) pooling equilibrium. For generic payoffs there will be only one pooling equilibrium payoff.

DEFINITION. A cheap-talk game has *partial common interests* if there exists a partition $J = \{J_i\}$, $i = 1, \ldots, j$, of T such that

$$\underline{v}_1(t_i; J_i) > \max \{v_1(t_i, a_l)\colon a_l \in \mathrm{BR}_2(J_l)\} \quad \text{for all } t_i \in J_i \text{ and } i \neq l, \tag{9}$$

$$\text{for all } t_i \in J_i \text{ there exists } a_i \in \mathrm{BR}_2(J_i; \pi) \text{ such that } v_1(t_i, a_i) > v_{1\mathrm{p}}(t_i), \tag{10}$$

and

$$\text{if } K \cap J_l \neq \phi \text{ for at least two } l, \text{ then for each } a \in \mathrm{BR}_2(K) \text{ there exist } i \text{ and } t_i \in K \cap J_i \text{ such that } \underline{v}_1(t_i; J_i) > v_1(t_i, a). \tag{11}$$

The definition is meant to capture the intuition that both players gain when types in sets J_i reveal (at least) that information. Condition (9) is a strong assumption guaranteeing that types in J_i prefer to identify themselves as members of J_i rather than as members of any other element of the partition. This condition alone implies the existence of an equilibrium in which the Sender communicates partial information. Condition (10) states that each type would prefer to identify herself as a member of the partition that contains it rather than be pooled. Conditions (9) and (10) guarantee that J is a self-signaling family of sets relative to any pooling equilibrium payoff.

Condition (11) requires that when members of different elements of the partition pool, at least one type loses relative to the least it could obtain when identified as a member of the set it belongs to. We use condition (11) to show that once the population arrives at a strategy that reveals the partition J, it will never drift to a less informative strategy profile. Condition (11) follows from (9) whenever $A = \bigcup_{i=1}^{j} \mathrm{BR}_2(J_i)$.

PROPOSITION 7. *If G is a costless signaling game with partial common interests, then no pooling equilibrium is an element of an ER set.*

If one accepts the interpretation that an evolutionary process cannot

return infinitely often to strategies that are not in any ER set, then Proposition 7 states that evolutionary pressures prevent agents in games with partial common interests from babbling uninformatively. The result does not require that an EES set exist (although it implies that the pooling equilibrium cannot be an element of an EES set).[8]

The Appendix contains a proof of Proposition 7. The basic idea of the argument is that, under our assumptions, the population can drift from a pooling equilibrium ρ to a strategy in which types separate according to the partition J, but once types have separated, they cannot drift back together.

We will describe the logic of the argument in a bit more detail. Let $R(\sigma)$ be the smallest closed set containing σ that satisfies (4). $R(\sigma)$ is the set of strategies that are reachable from σ via repeated invasions.

Now let J be the partition that appears in the definition of partial common interests. Define the set of separating strategies relative to J, Φ, as the set of strategies $\sigma = (\sigma_1, \sigma_2)$ such that

$$\text{if } \sigma_1(m, t_i) > 0 \text{ for } t_i \in J_i, \text{ then } \sigma_1(m, t) = 0 \quad \text{for all } t \notin J_i \tag{12}$$

and

$$\text{if } \sum_{t \in J_i} \sigma_1(m, t)\pi(t) > 0, \text{ then } \sigma_2(a, m) > 0 \quad \text{only if } a \in \mathrm{BR}_2(J_i). \tag{13}$$

Condition (12) states that types in different subsets of the partition use different messages. It places no restrictions on how types that are members of the same J_i signal: They may or may not all use the same message. It follows from (12) that the messages sent with positive probability by σ_1 can be partitioned into sets of messages sent by types in J_i for each i. Condition (13) states that the Receiver's response to any message sent by types in J_i is an optimal response to some conjecture (possiby not the correct one) supported on the types in J_i. The semipooling equilibrium outcome in which the Sender tells the Receiver which element of the partition contains t and the Receiver responds optimally is an element of Φ, but Φ typically contains other, nonequilibrium, strategies.

Once a strategy drifts into Φ it can never leave, and a pooling equilibrium strategy ρ must drift into Φ in the sense that $R(\rho) \cap \Phi \neq \phi$. The first claim follows from (11); if types in an element of the partition separate, then no strategy that pools them again can enter. The second claim follows

[8] It is possible to prove that the pooling equilibrium is not a member of an EES set if only (9) and (11) in the definition of partial common interests hold.

from (9) and (10). These observations provide the intuition for our result. A pooling outcome cannot be an element of an ER set because that set must reach the separating set, but once it does, it will never leave it.

The proposition does not assert that if the population begins at the pooling strategy it must inevitably drift to the separating set Φ. There may exist other ER sets that intersect $R(\rho)$. It only demonstrates that evolutionary pressures will not allow the population to return to the pooling outcome infinitely often.

In our definition of partial common interests, the pooling outcome plays a distinguished role. It is natural to ask whether we can use our arguments to show that an arbitrary equilibrium belongs to no ER set. The answer is yes, provided that we modify (10) and add another condition. Specifically, let $\sigma = (\sigma_1, \sigma_2)$ be an arbitrary equilibrium strategy. If there exists a partition $J = \{J_i\}$, $i = 1, \ldots, j$, of T such that (9), (11),

$$\text{for all } t_i \in J_i \text{ there exists } a_i \in \mathrm{BR}_2(J_i, \pi) \text{ such that } v_1(t_i, a_i) > \sum_{m \in M} \sum_{a \in A} v_1(t, a)\sigma_2(a, m)\sigma_1(m, t) \quad (14)$$

and

$$\text{for each } i, \text{ there exists an } m \text{ such that } J_i \subset \{t\colon \sigma_1(m, t) > 0\}, \quad (15)$$

then σ cannot be an element of any ER set. The same general argument that establishes Proposition 7 yields this result. Condition (14) plainly replaces (10); it states that all types of Sender prefer to be identified as members of the appropriate J_i rather than obtain the payoff from σ. Condition (15) holds automatically for the pooling outcome. The assumption guarantees that the partially pooling equilibrium relative to the partition J can invade the population strategy σ.

As an application, assume that the game has common interests and let J be the finest partition of T (each element of J contains a distinct element of T). Condition (14) holds for any inefficient equilibrium by the assumption of common interests; condition (15) holds automatically. Hence a result analogous to Proposition 1 holds for ER sets: In common-interest games with costless signaling if an equilibrium is an element of an ER set, then the equilibrium is efficient. A similar result holds for games with nominal signaling costs.

8. Related Work

In this paper we have discussed the impact that the possibility of innovation or invasion has on the strategies played in communication games. Our basic qualitative result is that these invasions tend to destabilize

	A	B	C	D
t_1	2,6	0,0	1,4	−10,1
t_2	0,0	2,6	1,4	−10,1

FIGURE 4

inefficient outcomes. Specifically, if the interests of the players coincide, then stable outcomes must be efficient. If the interests of the players agree to a lesser extent, partial communication must occur. In this section we present a detailed discussion of related literature, both to justify our approach and to put it and our results in perspective.

There have been a large number of ideas developed in order to describe what strategic behavior would persist in an environment subject to evolutionary pressures. The lack of consensus regarding what is the correct solution to evolutionary games suggests to us that there is no single correct solution; different methods will be useful in different contexts. In this section we discuss several possible alternative approaches. We are reassured that, to the extent that we have investigated them, other proposed solutions to evolutionary games provide results that are broadly similar to ours.

There are static variations of evolutionary stability that are not set valued. Wärneryd (1993) shows that every neutral ESS[9] in a cheap-talk game must be efficient in a game in which both players receive a positive payoff if the Receiver correctly guesses the Sender's type, and they receive zero otherwise. These pure-coordination games are games of common interest, so Wärneryd's results are consistent with the results in Section 4. Wärneryd's propositions do not extend to the broader class of common-interest games that we study. His solution concept does not permit strategies to drift freely off the equilibrium path; hence an inefficient outcome may be stable if there were an action that led to especially low payoffs (for example, less than the pooling equilibrium payoffs) for all types. For example, in the game depicted in Fig. 4, the strategy in which all Senders use the first message with probability one and Receivers respond to that message with action C, and respond with action D otherwise, is a neutral ESS. Selten's (1983) notion of limit ESS adds trembles to the game so that there are no unreached information sets. It appears that limit ESS's must exist and be efficient for the games Wärneryd studies, but not for

[9] 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).

	A	B	C	D
t_1	2,6	0,0	4,4	0,5
t_2	0,0	2,6	4,4	0,5

FIGURE 5

general common-interest games for the same reason that Wärneryd's results do not extend to such games.

There are also set-valued ideas that, like Swinkels's notion, attempt to describe limiting outcomes in population games subject to evolutionary forces. The notion of an evolutionarily stable (ES) set due to Thomas (1985a,b) weakens the entry condition in the definition of an EES set: Any strategy σ' that does better than the population strategy in a perturbed environment can enter; the entrant need not be an optimal response to the perturbed environment as in (4). For this reason, any ES set must contain an EES set. The arguments that we use to establish our destabilization results (Propositions 1, 3, and 6) also apply to ES sets. Existence results corresponding to Propositions 2, 4, and 5 would also hold for ES sets. Hence for cheap-talk games with common interests, ES sets exist and all elements of ES sets must be efficient (net of signaling costs). We choose to work with EES sets because ES sets are sensitive to the inclusion of dominated strategies. For example, in the game depicted in Fig. 5, the separating outcome is an element of an EES set. It fails to be an ES set because a strategy in which Senders use an unsent message and the Receiver responds to the invasion with the dominated action C can invade a population of Senders that separate.

Gilboa and Matsui's (1991) notion of cyclically stable sets (and the variants studied in Matsui (1991, 1992)) appear to be quite similar to our ER sets. The set of efficient outcomes in common-interest games must be a cyclically stable set. An efficiency result corresponding to Proposition 1 also appears to hold, at least for the coordination games studied by Wärneryd.

Although our approach is not explicitly dynamic, there is a connection between local stability of an outcome under the replicator dynamic (or closely related processes) and outcomes that satisfy static stability conditions. In particular, Cressman (1990) shows that in symmetric games Thomas's ES sets are necessarily locally stable for pure-strategy dynamics. The outcomes identified as evolutionarily stable in this paper, and only those outcomes, are locally stable with respect to the replicator dynamic.[10]

[10] In complementary work, Swinkels (1993) demonstrates that sets which are locally asymptotically stable relative to selection dynamics must contain a hyperstable (Kohlberg and Mertens, 1986) subset.

Canning (1992) and Nöldeke and Samuelson (1992) present adaptive dynamic models of games with communication. These papers prove that the only limiting outcomes are efficient in a subset of common-interest games with cheap talk.[11] (Canning assumes that the Sender and the Receiver have the same preferences over actions. Nöldeke and Samuelson study the same set of coordination games that Wärneryd analyzed.) The dynamics of these models allow the same kind of drift that occurs in EES sets.

The static evolutionary solution concept of Thomas (1985a,b) and the dynamics of Nöldeke and Samuelson (1992) permit actions at unreached information sets to drift out of the equilibrium component. As we noted in connection with our discussion of the game in Fig. 3, because EES limits invasions to strategies that respond optimally to the postentry environment, the population can only drift to strategies that support an equilibrium component. The distinction plays no role in common-interest games: The limited drift that we permit is sufficient to destabilize inefficient equilibria, while the efficient outcome is immune from all invasions. Whether or not the evolutionary process imposes discipline on actions taken at unreached information sets is ultimately a question about the underlying dynamics. We believe that our entry condition might arise in a dynamic model if entrants cannot change their strategy immediately following an invasion (but the existing population can).

Nöldeke and Samuelson also analyze Sender–Receiver games with two types of Sender and three actions for the Receiver. Their results for these games are consistent with the findings that we report in Section 5. In particular they show that the limiting outcome must be the Sender's favorite (when one exists).[12]

The clear advantage of the approach of Canning (1992) and Nöldeke and Samuelson (1992) is that it is explicitly dynamic. Therefore their results can be directly traced to behavioral assumptions on individual agents. These dynamic stories suggest realistic circumstances under which you must get cooperative outcomes in common-interest games. On the other hand, the approach necessarily leads to ergodic distributions so, at least at the level the models are currently applied, they allow no role for history in determining outcomes. These papers use much stronger

[11] Canning also shows that without mistakes his dynamic need not rule out inefficient outcomes as limit points even in pure coordination games.

[12] Nöldeke and Samuelson's result comes in the context of costless signaling games, in which our Example 2 illustrates that EES sets do not exist. We believe the difference arises because they assume that players do not change actions that respond optimally to their conjectures (even if alternative best responses exist). Therefore, players cannot drift from a pooling outcome to a payoff-equivalent equilibrium in which different types of Sender use different messages.

assumptions about the game than we do. Further work will determine whether our results continue to hold in a fully dynamic setting.

APPENDIX

Proof of proposition 1. Suppose that there exists $\sigma \in \Theta$ such that $U(\sigma) \neq u^*$. By Lemma 1 and repeated applications of Lemma 2 we can assume without loss of generality that there are distinct messages $m(t)$ for $t \in T$ such that

$$\sigma_1(m(t), t') = 0 \quad \text{for all } t \text{ and } t' \in T \tag{A1}$$

and

$$\sum_{a \in A} u_1(t, m(t), a)\sigma_2(a, m(t)) = \sum_{a \in A} \sum_{m \in M} u_1(t, m, a)\sigma_1(m, t)\sigma_2(a, m). \tag{A2}$$

In words, (A1) and (A2) state that message $m(t)$ is not used in σ, but the Sender of type t would not lose by sending $m(t)$. Define the strategy σ^* so that the Sender of type t uses message $m(t)$ with probability one, the Receiver responds optimally to messages $m(t)$ for $t \in T$, and the Receiver responds as σ_2 did to other messages. Since each Sender type uses a different message and the Receiver responds optimally to these messages, the Receiver must obtain his highest feasible payoff under σ^*; that is, $U_2(\sigma^*) = u_2^*$. Since the set of feasible payoffs of common-interest games has a unique Pareto-efficient point, it must be that σ^* is a Nash equilibrium and $U_1(\sigma^*) = u_1^*$. Further, σ^* is an optimal response to the population strategy σ, so it satisfies the invasion condition. It follows that for some $\varepsilon > 0$,

$$\sigma' \equiv \varepsilon\sigma^* + (1 - \varepsilon)\sigma \in \Theta. \tag{A3}$$

It is straightforward to check that if $U(\sigma) \neq u^*$, then $U(\sigma') \neq u^*$. Hence σ' cannot be a Nash equilibrium (at least one type of Sender t would do better by using the message $m(t)$ with probability one), which contradicts (A3) and completes the proof of the Proposition.

Proof of proposition 6. Assume, in order to obtain a contradiction, that there exists an EES set Θ that contains σ. Let $\overline{m}_i$ denote a message used by the types of J_i ($\sigma_1(\overline{m}_i, t) = 1$ for all $t \in J_i$). By Lemmas 1 and 2 we can assume, without loss of generality, that there exists a message

m_i' such that $\sigma_1(m_i', t) = 0$ for all $t \in T$ and $\sigma_2(a, m_i') = \sigma_2(a, \overline{m}_i)$ for all a. Consider the strategy $\sigma' = (\sigma_1', \sigma_2')$, where, letting a_i^* denote the Receiver's optimal response to J_i as in the definition of self signaling,

$$\sigma_1'(m, t) = \begin{cases} \sigma_1(m, t) & \text{for } t \notin J \\ 1 & \text{for } t \in J_i \text{ and } m = m_i' \\ 0 & \text{for } t \in J_i \text{ and } m \neq m_i' \end{cases}$$

and

$$\sigma_2'(m, t) = \begin{cases} \sigma_2(a, m) & \text{for } m \neq m_i' \quad \text{for all } i \\ a_i^* & \text{for } m = m_i' \end{cases}.$$

Since σ_1' is an optimal response to both σ_2 and σ_2', it is an optimal response to all mixtures. Since σ_2' is an optimal response to σ_1, it is an optimal response to $\varepsilon\sigma_1' + (1 - \varepsilon)\sigma_1$ for some $\varepsilon > 0$ provided that payoffs are generic. It follows that $\varepsilon\sigma' + (1 - \varepsilon)\sigma \in \Theta$ for ε sufficiently small. However, since it is in the interest of all types in J_i to use m_i' instead of $\overline{m}_i$, $\varepsilon\sigma' + (1 - \varepsilon)\sigma$ fails to be a Nash equilibrium when $\varepsilon > 0$. This observation contradicts the definition of EES sets and completes the proof.

Proof of proposition 7. For any strategy profile σ, $R(\sigma)$ is the smallest closed set containing σ that satisfies (4). Since the intersection of any two sets that satisfy (4) is either empty or satisfies (4) and is closed, $R(\cdot)$ is defined for all strategies. We claim that

$$\text{if there exists an ER set } \Theta \text{ such that } \sigma \in \Theta, \text{ then } R(\sigma) = \Theta. \quad \text{(A4)}$$

In order to prove (A4) note that Θ must contain $R(\sigma)$ because Θ satisfies (4), and by minimality the sets must be equal.

Next we show

$$\text{if } \rho \text{ is a pooling equilibrium, then } R(\rho) \cap \Phi \neq \phi \quad \text{(A5)}$$

and

$$\text{if } \sigma \in \Phi, \text{ then } R(\sigma) \subset \Phi. \quad \text{(A6)}$$

Condition (A5) states that the set of strategies reachable via invasions from a pooling equilibrium must intersect the separating set Φ. To prove (A5), first observe that by the arguments of Lemmas 1 and 2 we may

assume without loss of generality that there are j unused messages (one for each member of the partition J), and the Receiver responds to these messages as he responds to messages on the equilibrium path. By the definition of partial common interest, we know that there exists a partially pooling equilibrium outcome relative to the partition J. Let σ' be an equilibrium that supports the partially pooling outcome in which the Sender only signals with the unused messages of ρ and the Receiver responds to unsent signals with the pooling action of ρ_2. It follows from (4) and the definition of $R(\rho)$ that there exists $\varepsilon' > 0$ such that for all $\varepsilon \in (0, \varepsilon')$, $(1 - \varepsilon)\rho + \varepsilon\sigma' \in R(\rho)$. In fact, since σ' can enter when the population strategy is of the form $(1 - \varepsilon)\rho + \varepsilon\sigma'$ for any $\varepsilon \in (0, 1)$, it follows that $\sigma' \in R(\rho)$.

Condition (A6) states that once the population enters the separating set it cannot drift out of it. To establish (A6), let $\sigma \in \Phi$, and suppose that there exists σ' that satisfies (4). We must show that $\sigma' \in \Phi$. First we will show that types in different elements of the partition use different messages under σ_1'. By (9) it follows that if $t_i \in J_i$, then $\sigma_1'(m, t_i) = 0$ whenever $\sigma_1(m, t) > 0$ for $t \notin J_i$, so if members of different elements of the partition choose the same message under σ_1', then it must be a message that is not sent under σ_1. Let m be such that $\sigma_1(m, t) = 0$ for all t and assume, in order to reach a contradiction, that $K \cap J_l \neq \phi$ for at least two l, where $K = \{t\colon \sigma_1'(m, t) > 0\}$. It follows from (4) that $\sigma_2'(\cdot, m)$ must put positive weight only on strategies in $\mathrm{BR}_2(K)$. Therefore, from (11) and the definition of Φ, at least one type in K would do better following σ_1 than playing according to σ_1'. This contradicts (4), so it must be that types in different elements of the partition use different messages under σ_1'. In order for σ_2' to satisfy (4), it must respond optimally to σ_1 following all messages that σ_1 uses with positive probability, and it must respond optimally to σ_1' following all messages used with positive probability only by σ_1'. Therefore (13) must hold and so $\sigma' \in \Phi$.

We now prove the proposition. In order to obtain a contradiction, assume that there exists an ER set Θ such that $\rho \in \Theta$. From (A4) it follows that $R(\rho) = \Theta$. Therefore, there exists $\sigma \in R(\rho) \cap \Phi$ by (A5). It follows from (A4) that

$$R(\sigma) = \Theta \tag{A7}$$

and from (A6) that

$$R(\sigma) \subset \Phi. \tag{A8}$$

(A7) and (A8) are not consistent since $\rho \in \Theta$ but $\rho \notin \Phi$.

REFERENCES

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

BLUME, A. (1992). "Equilibrium Refinements in Perturbed Games and in Sender-Receiver Games." Iowa City: University of Iowa.

BLUME, A., AND SOBEL, J. (1991). "Communication-Proof Equilibrium in Cheap-Talk Games." La Jolla: UCSD.

CANNING, D. (1992). "Learning Language Conventions in Common Interest Signaling Games." New York: Columbia University.

CRESSMAN, R. (1990). "Evolutionarily Stable Sets in Symmetric Extensive Two-Person Games."

FARRELL, J. (1993). "Meaning and Credibility in Cheap-Talk Games," *Games Econ. Behav.*, **5,** 514–531.

FUDENBERG, D., AND MASKIN, E. (1990). "Evolution and Cooperation in Noisy Repeated Games," *Amer. Econ. Rev.* **80,** 274–279.

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

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

HOFBAUER, J., AND SIGMUND, K. (1988). *The Theory of Evolution and Dynamical Systems.* Cambridge: Cambridge Univ. Press.

KALAI, E., AND SAMET, D. (1984). "Persistent Equilibria in Strategic Games," *Int. J. Game Theory* **13,** 129–144.

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

KOHLBERG, E., AND MERTENS, J.-F. (1986). "On the Strategic Stability of Equilibria," *Econometrica* **54,** 1003–1038.

MATSUI, A. (1991). "Cheap-Talk and Cooperation in a Society," *J. Econ. Theory* **54,** 245–258.

MATSUI, A. (1992). "Best Response Dynamics and Socially Stable Strategy," *J. Econ. Theory* **57,** 343–362.

MATTHEWS, S., OKUNO-FUJIWARA, M., AND POSTLEWAITE A. (1991). "Refining Cheap-Talk Equilibria," *J. Econ. Theory* **55,** 247–273.

MAYNARD SMITH, J. (1982). *Evolution and the Theory of Games.* New York: Cambridge Univ. Press.

MAYNARD SMITH, J., AND PRICE G. (1973). "The Logic of Animal Conflict," *Nature* **246,** 15–18.

MYERSON, R. (1988). "Credible Negotiation Statements and Coherent Plans," *J. Econ. Theory* **48,** 264–291.

NÖLDEKE, G., AND SAMUELSON, L. (1992). "The Evolutionary Foundations of Backward and Forward Induction," Bonn and Wisconsin Discussion Paper.

NÖLDEKE, G., AND SAMUELSON, L. (1993). "An Evolutionary Analysis of Backward and Forward Induction," *Games Econ. Behav.*, **5,** 425–454.

RABIN, M. (1990). "Communication Between Rational Agents," *J. Econ. Theory* **51,** 144–170.

SELTEN, R. (1983). "Evolutionary Stability in Extensive Two-Person Games," *Math. Soc. Sci.* **5,** 269–363.

SWINKELS, J. (1992a). "Evolutionary Stability with Equilibrium Entrants," *J. Econ. Theory* **57,** 306–332.

SWINKELS, J. (1992b). "Stability and Evolutionary Stability," *J. Econ. Theory* **57,** 333–342.

SWINKELS, J. (1993). "Adjustment Dynamics and Rational Play in Games," *Games Econ. Behav.* **5,** 455–484.

THOMAS, B. (1985a). "On Evolutionarily Stable Sets," *J. Math. Biology* **22,** 105–115.

THOMAS, B. (1985b). "Evolutionarily Stable Sets in Mixed-Strategist Models," *Theoretical Population Biology* **28,** 332–341.

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

WÄRNERYD, K. (1993). "Cheap Talk, Coordination, and Evolutionary Stability," *Games Econ. Behav.*, **5,** 532–546.

ZAPATER, I. (1991). "Generalized Communication Between Rational Agents." Providence: Brown University.