Evolutionary Stability with Equilibrium Entrants*

JEROEN M. SWINKELS

Department of Economics, Stanford University, Stanford, California 94305-6072

Received November 2, 1990; revised January 11, 1992

The strategy that upsets a potential evolutionarily stable strategy may in itself be very unstable, or may differ from the candidate strategy only in irrelevant ways. This paper develops a solution concept addressing these difficulties. We look for a set of Nash equilibria such that small groups of entrants whose members are satisfied with their entry cannot take the population out of the set. Such a set is robust to the iterated removal of weakly dominated strategies, depends only on the reduced normal form, and has the never a weak best response property. For generic two person extensive form games, such sets generate payoffs consistent with proper equilibria. Journal of Economic Literature Classification Number: C72 © 1992 Academic Press, Inc.

I. INTRODUCTION AND MOTIVATION

When pairs of players from a single population are randomly, repeatedly, and anonymously paired to play a particular two person symmetric game, evolutionary stability [8, 9] provides a justification for equilibrium starkly different from those advanced in traditional game theory. Whereas traditional game theory tests the rationality (and a great deal more) of a strategy, evolutionary stability tests the stability of a population playing a given strategy to small groups of mutants playing other strategies. A status quo strategy is evolutionarily stable if, for every differing strategy for the entrants, the status quo strategy does better on average when matched against a random member of a population made up of almost all status quo players and a few entrants than does the entrant.

While evolutionary stability does not explain how a population gets to such a strategy, it seems a natural test of the stability of such a strategy

* I thank the Olin Foundation and the Sloan Foundation for financial support during the writing of this paper. I also thank the many people who made helpful comments on this work, including Abhijit Banerjee, Eric van Damme, Avinash Dixit, David Kreps, George Mailath, Akihiko Matsui, Paul Milgrom, Barry Nalebuff, Michele Piccione, Larry Samuelson, Christina Shannon, Joel Sobel, Jörgen Weibull, an anonymous referee, and participants in seminars at Princeton, Stanford, U.B.C., Wisconsin, the University of Pennsylvania, Northwestern, and Cornell.

0022-053l/92 $5.00
Copyright © 1992 by Academic Press, Inc.
All rights of reproduction in any form reserved.
once reached. Any small mutant group does worse than the status quo and so is selected against.

Evolutionary stability also can provide an attractive basis for equilibrium in economic environments. Our model here might be one of experimentation and imitation. Strategies that do well are imitated, and thus are played by an increasing fraction of the population. Strategies that do less well are eventually dropped in favor of more profitable strategies. Evolutionary stability is the condition that, given what the actors in the economy are doing, any small entrant group that tries something different will not do as well as members of the status quo in the post-entry environment. There is thus no incentive for members of the status quo population to imitate the experimenters, but a positive incentive for the experimenters to join the status quo.

A major attraction of evolutionary stability is that it implies behavior that is consistent with some very strong ideas of rationality. Thus for example, van Damme [16, Theorem 9.3.4] shows that an evolutionarily stable strategy is proper [10].

Evolutionary stability makes no restrictions on the strategies played by entrants. As a result, the strategy for the entrants that upsets a potential evolutionarily stable strategy (ESS) can itself be very unstable. This has two implications. First, most games do not have an ESS. Second, if some of the deviations against which evolutionarily stable strategies are tested are implausible, then results such as van Damme's on the implications of evolutionary stability become less compelling. If these implications fail to hold true when entrants are restricted to a more reasonable set, then evolutionary stability loses much of its attraction.

We argue that in economic environments it may be reasonable to restrict attention to entrant strategies that themselves satisfy a weak optimality/stability requirement. In particular, we examine the implications of assuming that entering groups choose best responses to their environments. Thus, if the status quo strategy profile is \(\sigma\), and entrants make up a proportion \(\varepsilon\) of the population, then we only consider entering strategies \(\mu\) that are a best response to a population a proportion \(1 - \varepsilon\) of which is playing \(\sigma\) and a proportion \(\varepsilon\) of which is playing \(\mu\). We refer to such a \(\mu\) as an equilibrium entrant: the entrants are in an equilibrium of the game generated by the original game and the actions of the incumbent players.

This restriction on entrants can be supported in a variety of ways. The first is to relax the assumption of evolutionary stability that entrants have no rationality whatsoever. Endowing entering (experimenting) players with the foresight to choose best responses (to the post-entry environment) strikes a middle ground between the total absence of rationality assumed by evolutionary game theory and the strenuous rationality requirements of traditional game theory.
The basic behavioral assumption behind the second justification is that incumbents do not instantly respond to every small change in their environment. Rather, they wait until they are convinced of the direction and the permanence of the change. Thus, even if an original entry $\mu$ is not a best response to $(1 - \varepsilon) \sigma + \varepsilon \mu$, incumbents may not respond until the entrants have adjusted among themselves.

This seems a pretty realistic observation about the way in which economic agents actually make decisions. It is also easily supportable as rational. If there is any cost to changing strategies, then incumbents, before incurring the cost of responding to an entrant group, might well ask themselves whether the innovation will be around for any length of time. If the judgment of the incumbents is that the innovation will either die out or change significantly in the near future, then, if change is at all costly, it will not pay incumbents to respond. Given this, if the entrants are indeed playing a nonoptimal strategy, they will be under pressure to change, justifying the expectation of the incumbents that the initial entering strategy was transient. Conversely, if the entrants are playing a strategy $\mu$ that is a best response to $(1 - \varepsilon) \sigma + \varepsilon \mu$, then there is no positive incentive for them to change, and thus if $\sigma$ is not a best response to $(1 - \varepsilon) \sigma + \varepsilon \mu$ then it seems reasonable to expect the incumbents eventually to react.

A third justification occurs when the population is composed of players who differ in their ability to change strategies. To take an extreme case, assume that most of the population can choose a strategy only once, and then must play it forever. The balance of the population can costlessly change strategies at any time. In these circumstances, it can never be the case that the fast movers, acting optimally, will persist at a strategy choice $\mu$ that was not a best response to $(1 - \varepsilon) \sigma + \varepsilon \mu$. Testing the robustness of a population to such entrants is then a test of whether the players who are unable to change their strategies are at a disadvantage in this environment.

A final justification of the test on entrants envisages a world with coordinated decisions by entering players to play a deviating strategy. After all, if the population as a whole is playing a Nash equilibrium, then unilateral deviations cannot help a player. One might thus think of a deviation from the status quo as occurring only when a group of players has collectively decided to make that deviation. The condition on entrants is then a minimal test of the stability of the agreement among the players who deviated. Members of any deviating group that did not satisfy the condition would have an immediate incentive to deviate from the agreed upon deviation.\(^1\)

\(^1\) This is in tension with the random matching assumption, but not logically contradictory to it. Consider an auto club the members of which decide to follow some new driving practice. Members cannot identify each other on the road. The entry restriction then precisely asks that members of the auto club who assume that other members will abide by the decision will not find it in their interest to deviate.
Of course, some of these justifications may imply further restrictions on entrants. For example, if the equilibrium entry condition is a result of entrants learning about their environment in the face of slow to change incumbents (or incumbents unable to change their strategy), then, depending on the theory of learning one finds attractive, various stronger (or weaker) conditions on entrants might arise. Similarly, if entry playing a deviating strategy is a coordinated action by a group of players, then one also might examine the incentives for subgroups of the deviating group to deviate further, in the spirit of Bernheim, Peleg, and Whinston [1]. In the companion paper we show that much of the analysis survives a restriction similar to perfection [14] on entrants.

We begin by modifying evolutionary stability to reflect the restriction to equilibrium entrants. We look for symmetric strategy profiles of symmetric games (not necessarily two player) that are robust against symmetric equilibrium entrants. The resultant concept, robustness against symmetric equilibrium entrants, is weaker than evolutionary stability but retains many of its more attractive and surprising implications. For example, a strategy that is robust against symmetric equilibrium entrants is always proper.

We then turn to asymmetric games (what Selten [12] refers to as truly asymmetric contests). Applying the idea of (asymmetric) equilibrium entrants to asymmetric games is simpler than the standard evolutionary approach to asymmetric games and yields existence on a broader class of games.

An ESS may fail to exist for another reason. Evolutionary stability is often unable to differentiate between strategies that do not differ in equilibrium. In asymmetric games for instance, a strategy σ will be an ESS only if there is no strategy other than σ that is a best response to σ. Games that have such alternate best responses include most games with an interesting extensive form. Restricting attention to equilibrium entrants does not solve this difficulty.

This leads us to consider a solution concept that maintains the same restriction on entrants, but allows for a set valued solution. Informally, a set of Nash equilibria is equilibrium evolutionarily stable (EES) if equilibrium entrants cannot take the population out of the set. Because we restrict attention to sets of Nash equilibria, if the average strategy in a status quo population is in an EES set, and entrants are optimal, then no member of the status quo population will ever face a positive incentive to change strategies. Thus, even if current entrants anticipate that there will be future entrant groups, choosing a best response paying attention only to the population configuration immediately following their own entry is an optimal entry strategy, as long as other players are following the same or a more restrictive entry strategy.

EES sets have a particularly nice interpretation for normal form games.
that "come from" a generic extensive form game. In such a situation, an EES set corresponds to a single outcome in the extensive form. The size of an EES set thus simply reflects that there is a range of behavior at out of equilibrium information sets that will support the required behavior at reached information sets. The test of an EES set is then roughly the following: if behavior at out of equilibrium information sets drifts far enough to give players a positive incentive to deviate from the equilibrium path, and if they do so in an optimal manner, will the newly created incentives validate the direction of drift, or will those players who drifted wish they had done something else?

This test is very different from the considerations that sequential equilibrium uses to drive actions at out of equilibrium information sets. Nonetheless, it turns out that EES sets for two player generic extensive form games always contain a proper element. Thus [18] the outcome supportable by an EES set will be sequential in the associated generic extensive form.

In fact, EES sets satisfy considerably stronger restrictions than do sequential equilibria. The iterative removal in any order of weakly dominated strategies will never completely eliminate an EES set. EES sets also satisfy Kohlberg and Mertens' [5] never a weak best response property. The outcome in a generic extensive form associated with an EES thus will for example satisfy the Intuitive Criterion (Cho and Kreps [2]). Under some additional conditions EES sets generate payoffs consistent with stable equilibria [4, 5]; see the companion paper [15] for details.

Matsui [7] has shown a close connection between our concept of robustness against equilibrium entrants and Gilboa and Matsui's [3] cyclically stable set. Gilboa and Matsui examine populations in which the population strategy profile can evolve in any direction that is a best response to the current population profile. A cyclically stable set is a set of strategy profiles that is minimal with respect to being closed under this process (see [3] for a formal development). Matsui modifies this process, requiring a piecewise linear path. He shows that a set of strategy profiles (not necessarily Nash equilibria) will be closed under the modified dynamic if and only if it is robust to equilibrium entrants. This goes some distance toward providing a dynamic rational for equilibrium evolutionary stability.2

2It should be noted that while a population beginning within a set robust against equilibrium entrants can never leave that set under the modified Gilboa–Matsui dynamic, a population beginning arbitrarily close to the set can cycle away from the set. An example is the game of Fig.9.3.1 in [16]. Also, as is explained following Definition 10, I am a little dubious about the plausibility of myopic decision procedures when they can yield consistently suboptimal choices. This is true of the Gilboa–Matsui dynamic for games where cyclically stable sets are not subsets of Nash equilibria.
Section II deals with some preliminaries. Section III covers the existing definition of evolutionary stability. Section IV presents a simple example in which an unstable entrant upsets a potential ESS. Section V introduces robustness against equilibrium entrants. Section VI considers asymmetric games. Section VII examines the difficulties of robustness in games with alternate best responses to a candidate strategy. Section VIII presents the set valued solution concept, equilibrium evolutionary stability, and explores its properties. A reading of Section II, followed by examination of the examples in Sections IV and VII, and then a full reading of Section VIII will cover equilibrium evolutionary stability itself. Much of the rest of the paper examines the connection between equilibrium evolutionary stability and evolutionary stability.

II. Preliminaries

We restrict attention to finite normal form games. Players \(i = 1, \ldots, n\) have finite pure strategy sets \(S_i\) with \(S = \prod_{i=1}^{n} S_i\). The payoff function is \(\pi = (\pi_1, \ldots, \pi_n)\). The game generated by \(S\) and \(\pi\) is denoted \((S, \pi)\). The space of mixed strategies is \(\Phi = \prod_{i=1}^{n} \Delta(S_i)\). I often fail to differentiate between the pure strategy \(s_i\), and the mixed strategy for \(i\) that places probability 1 on \(s_i\). Typical elements of \(\Phi\) are denoted by lower case greek letters. The vector of weights given by the mixed strategy profile \(\sigma = (\sigma_1, \ldots, \sigma_n)\) to \(s \in S\) is \(\sigma(s) = (\sigma_1(s_1), \ldots, \sigma_n(s_n))\). The mixed strategy profile that differs from \(\sigma\) only in that player \(i\) plays \(\gamma_i\) instead of \(\sigma_i\) is denoted \(\sigma \setminus \gamma_i\). \(\pi\) is extended to \(\Phi\) in the standard way. The carrier of \(\eta\) is \(C(\eta) = (C_1(\eta_1), \ldots, C_n(\eta_n));\) i.e., \(C_i(\eta_i)\) is the set of pure strategies played with positive probability by \(i\). The best response correspondence into pure strategies is \(B(\eta) = (B_1(\eta), \ldots, B_n(\eta))\), where \(B_i(\eta) = \{s_i \in S_i | \pi_i(\eta \setminus s_i) \geq \pi_i(\eta \setminus t_i) \forall t_i \in S_i\}\). For \(X_i \subseteq S_i\), \(i = 1, \ldots, n\), and \(X = \prod_{i=1}^{n} X_i\), \((X, \pi)\) denotes the game generated from \((S, \pi)\) by restricting strategy choices to those having carrier \(X\). The set of best responses over a subset \(X = \prod_{i=1}^{n} X_i\) is \(B_X(\eta) = (B_{X_1}(\eta), \ldots, B_{X_n}(\eta))\), where \(B_{X_i}(\eta) = \{s_i \in X_i | \pi_i(\eta \setminus s_i) \geq \pi_i(\eta \setminus t_i) \forall t_i \in X_i\}\). The set of Nash equilibria of \((S, \pi)\) is \(N(S, \pi)\). \(D(\mu, \nu)\) is the Euclidean distance between \(\mu, \nu \in \Phi\). For \(Y\) a closed subset of \(\Phi\), and \(\mu \in \Phi\), \(D(\mu, Y) = \min_{\nu \in Y} D(\mu, \nu)\).

III. Evolutionarily Stable Strategies

The concept of an evolutionarily stable strategy (ESS) was formalized by Maynard Smith [8, 9]. Two insights underlie evolutionary stability. First,
many biological situations can be understood as games for which organisms have genetically programmed strategies and for which the payoffs reflect reproductive fitness. Second, evolutionary forces will tend to give these strategies equilibrium properties. An evolutionarily stable strategy is one that, when played by almost all the population, does better than any mutant strategy. The model is of a symmetric two player game, with meetings within a single population. It is thus appropriate to restrict attention to symmetric strategies. Formally:

**Definition 1.** For a two person symmetric game \((S, \pi)\), a symmetric strategy profile \(\sigma\) is an *evolutionarily stable strategy* (ESS) if \(\exists \epsilon > 0\) such that \(\forall \epsilon \in (0, \epsilon')\), and for any symmetric \(\mu\), \(\pi_1(\mu_1, (1-\epsilon)\sigma_2 + \epsilon\mu_2) \geq \pi_1(\sigma_1, (1-\epsilon)\sigma_2 + \epsilon\mu_2) \Rightarrow \mu = \sigma\).

Thus, for each symmetric strategy profile \(\mu \neq \sigma\), there is an \(\epsilon'\) such that if players playing \(\mu\) make up less than \(\epsilon'\) of the population, then they will do worse against the population than those playing \(\sigma\), and thus presumably will be driven out by evolutionary pressure.

An equivalent definition of an ESS is that

1. \(\sigma\) is a symmetric Nash equilibrium of \((S, \pi)\), and
2. for any symmetric \(\mu = (\mu_1, \mu_2), \mu \neq \sigma\), such that \(\pi_1(\mu_1, \sigma_2) = \pi_1(\sigma_1, \sigma_2), \pi_1(\mu_1, \mu_2) < \pi_1(\sigma_1, \mu_2)\).

This is the standard formulation. The definition given will facilitate comparison with later definitions.

Van Damme [16] gives a good exposition of the properties of evolutionary stability. Among the more surprising properties of evolutionary stability is its relation to Myerson’s proper equilibrium [10]:

**Definition 2.** For \(\epsilon > 0\), the strategy profile \(\sigma^\epsilon\) is an \(\epsilon\)-proper equilibrium of \((S, \pi)\) if \(\sigma^\epsilon\) is completely mixed and if \(\forall i \in 1, \ldots, n \forall r_i, t_i \in S_i, \pi_i(\sigma^\epsilon \setminus r_i) < \pi_i(\sigma^\epsilon \setminus t_i) \Rightarrow \sigma^\epsilon_i(r_i) \leq \epsilon \sigma^\epsilon_i(t_i)\).

**Definition 3.** The strategy profile \(\sigma\) is a *proper equilibrium* of \((S, \pi)\) if there exists a sequence \(\{\sigma^\epsilon\}_{\epsilon \downarrow 0}\), \(\sigma^\epsilon \to \sigma\) such that \(\forall \epsilon, \sigma^\epsilon\) is an \(\epsilon\)-proper equilibrium of \((S, \pi)\).

Van Damme [16] shows that if \(\sigma\) is an ESS for \((S, \pi)\), then \(\sigma\) is a proper equilibrium for \((S, \pi)\). We generalize this result in Theorem 5.

**IV. An Example**

For many seemingly “stable” games, an ESS fails to exist. Consider the game of Fig. 1.
Either player can veto the subgame; payoffs in the event that the subgame is not reached do not depend on which player(s) chose to make the subgame unreachable. The subgame itself has a prisoner's dilemma structure. Given that \((D, D)\) in the subgame gives each player less than the value of the outside option, it seems reasonable that both players will avoid it.

Figure 2 gives the reduced normal form\(^3\) of this game. The unique symmetric Nash equilibrium, in which I and II choose \(O\), is not evolutionarily stable: a viable entering strategy is \(\text{InC}\). Considered in terms of the underlying extensive form, the reason \(O\) fails is that a group of deviants who choose \(\text{In}\) followed by \(C\) in the subgame can enter the population. When a deviant meets a status quo player, the subgame is not reached, so the poor strategy choice for the subgame does not hurt. When two deviants meet, there is nothing in the concept of evolutionary stability to rule out cooperation in the subgame. However, because \(C\) is strictly dominated in the subgame, it seems unlikely that this behavior by the deviants will persist. Thus, the fact that \(O\) is not robust against \(\text{InC}\) need not imply the "instability" of \(O\).

\(^3\)The reduced normal form of a game is obtained by eliminating redundant pure strategies. A full discussion of the reduced normal form can be found in [16].
V. ROBUSTNESS AGAINST SYMMETRIC EQUILIBRIUM ENTRANTS

The fact that a candidate equilibrium may fail to be evolutionarily stable because it is non-robust to an entrant that is itself unstable, as in the previous example, motivates our first modification of the definition of evolutionary stability. Instead of requiring a strategy to be robust against all possible entrants, we require robustness only against entrants that are themselves best responses to their environment:

**Definition 4.** For a symmetric game \((S, \pi)\), a symmetric strategy profile \(o\) is **robust against symmetric equilibrium entrants (RSEE)** if there exists \(\epsilon' > 0\) such that \(\forall \epsilon \in (0, \epsilon')\), and for any symmetric \(\mu\), \(C(\mu) \subseteq B((1 - \epsilon) \sigma + \epsilon \mu) \Rightarrow \mu = \sigma\).

Evolutionary stability is the condition that no strategy \(\mu \neq \sigma\) is at least as good a response as is \(\sigma\) against a population that is almost all \(\sigma\) and a little \(\mu\). Robustness is the condition that no strategy \(\mu \neq \sigma\) is a best response against a population that is almost all \(\sigma\) and a little \(\mu\). For the game of Fig. 2, \((0, 0)\) is RSEE but not an ESS. Note that the restriction to two player games has been dropped.

A condition equivalent to \(\sigma\) being RSEE is

1. \(\sigma\) is a symmetric Nash equilibrium of \((S, \pi)\), and
2. \(\sigma\) is the only symmetric Nash Equilibrium of \((B(\sigma), \pi)\).

To see the equivalence, note first that for any \(\epsilon \in (0, 1]\), \(B((1 - \epsilon) \sigma + \epsilon \cdot )\) inherits the relevant properties of \(B(\cdot)\), and so there is (a symmetric) \(\mu\) with \(C(\mu) \subseteq B((1 - \epsilon) \sigma + \mu)\). Since \(\sigma\) is RSEE, \(\sigma = \mu\) and so \(C(\sigma) \subseteq B(\sigma)\), i.e., \(\sigma\) is Nash. Since \(C(\sigma) \subseteq B(\sigma)\), for \(\epsilon\) sufficiently small, \(B((1 - \epsilon) \sigma + \epsilon \mu) \subseteq B(\sigma)\). Thus, \(C(\mu) \subseteq B((1 - \epsilon) \sigma + \epsilon \mu) \Rightarrow C(\mu) \subseteq B_{B(\sigma)}(\mu) \Rightarrow \mu \in N(B(\sigma), \pi)\).

**Observation 1.** If \(\sigma\) is an ESS of \((S, \pi)\), then \(\sigma\) is RSEE in \((S, \pi)\).

While weaker than evolutionary stability, robustness does share some of evolutionary stability's properties. In particular, if \(\sigma\) is an RSEE profile for \((S, \pi)\), then \(\sigma\) is a proper equilibrium for \((S, \pi)\). This is again an implication of Theorem 5 below.

We also have:

**Theorem 1.** For a game \((S, \pi)\), let \(S'\) be obtained from \(S\) by the removal of the same weakly dominated strategy for each player. If \(\sigma\) is RSEE in \((S, \pi)\) then \(C(\sigma) \subseteq S'\), and \(\sigma\) is also RSEE in \((S', \pi)\).

**Proof.** Consider \((B_S(\sigma), \pi)\). Any Nash equilibrium of \((B_S(\sigma), \pi)\) is also a Nash equilibrium of \((B_S(\sigma), \pi)\). Thus for \(\sigma\) to be the unique element of \(N(B_S(\sigma), \pi)\), it must be the case that \(C(\sigma) \subseteq S'\) and that \(\sigma\) is the unique element of \(N(B_S(\sigma), \pi)\).
Thus, if $\sigma$ is RSEE in $(S, \pi)$, then $\sigma$ also is RSEE for any game attained from $(S, \pi)$ by the symmetric iterative removal in any order of weakly dominated strategies. It is also clear that if $\sigma$ is RSEE in $(S, \pi)$, and $(S''', \pi''')$ is obtained from $(S, \pi)$ by the symmetric addition of a strongly dominated strategy, then $\sigma$ is also RSEE in $(S''', \pi''')$.

VI. ASYMMETRIC GAMES

In this section, we display the standard approach in biological game theory to asymmetric games. We show that this approach is equivalent to a simpler approach, but that either approach yields existence only on a very limited class of asymmetric games. We also show that the natural generalization of robustness to asymmetric games broadens the class of asymmetric games for which equilibrium is defined, and has similar properties to RSEE.

The standard approach to asymmetric games in biological game theory is to symmetrize the game by having nature assign player roles (see [12]). Thus, for the asymmetric two person game $(S, \pi)$, it is assumed that each individual in the population has an equal chance of being called upon to move as player 1 or player 2. A strategy is thus a specification of what to do as player 1, and what to do as player 2. Thus, each player's pure strategy space is $T = \{(s_1, s_2) | s_1 \in S_1$ and $s_2 \in S_2\}$. The payoff $\Pi$ to $((s_1, s_2), (t_1, t_2)) \in T$ is thus given by $2\Pi_1 ((s_1, s_2), (t_1, t_2)) = \pi_1 (s_1, t_2) + \pi_2 (t_1, s_2)$, and $2\Pi_2 ((s_1, s_2), (t_1, t_2)) = \pi_1 (t_1, s_2) + \pi_2 (s_1, t_1)$. It is easily verified that $(T^2, \Pi)$ is a symmetric game. A strategy profile $(\sigma_1, \sigma_2) \in \Phi$ is said to be evolutionarily stable in $(S, \pi)$ if $((\sigma_1, \sigma_2), (\sigma_1, \sigma_2))$ is evolutionarily stable in $(T^2, \Pi)$.

A more straightforward approach might be to retain the asymmetry of the game, but allow asymmetric entrants also. Allowing asymmetric entrants also seems appropriate if the game is symmetric, but the players for the different positions are drawn from separate populations.

**Definition 5.** The strategy profile $\sigma$ is an asymmetric evolutionarily stable strategy (AESS) for the game $(S, \pi)$ if $\exists \epsilon > 0$ such that $\forall \epsilon \in (0, \epsilon')$ and $\forall \mu,$

$$\pi_i (((1 - \epsilon) \sigma + \epsilon \mu) \setminus \mu_i) \geq \pi_i (((1 - \epsilon) \sigma + \epsilon \mu) \setminus \sigma_i) \forall i \Rightarrow \mu = \sigma.$$

$(\sigma_1, \sigma_2)$ mixes over $S_1$ independently of $S_2$. Although not every mixture over $T$ can be expressed as such an independent mixture over $S_1$ and $S_2$, every mixture over $T$ is payoff equivalent to such a mixture. This is because such a strategy is effectively a behavior strategy of a game in which nature first selects which player fills which role. For our purposes, it is enough to consider such mixtures.
Thus there is an $\varepsilon' > 0$ such that for every $\mu \neq \sigma$, if players playing $\mu$ make up less than $\varepsilon'$ of the population, then for at least one $i$, $\mu_i$ does worse against the population than does $\sigma_i$.

For two person games, this definition is equivalent to the conventional method of extending evolutionary stability to asymmetric games. Thus, both methods have very poor existence properties. Formally:

**Definition 6** The strategy profile $\sigma$ is strict in the game $(S, \pi)$ if $\exists s \in S$ such that $\sigma(s) = 1$ and $\{s\} = B(\sigma)$.

**Theorem 2.** For two player games, $\sigma$ is an AESS of $(S, \pi) \Rightarrow (\sigma, \sigma)$ is an ESS of $(T^2, \Pi)$. For games with any finite number of players, $\sigma$ is an AESS of $(S, \pi) \Rightarrow \sigma$ is strict in $(S, \pi)$.

**Proof.** We prove that $(\sigma, \sigma)$ an ESS of $(T^2, \Pi)$ is equivalent to $\sigma$ strict in $(S, \pi)$ for two player games, and that AESS is equivalent to strictness for $n$ player games.

(Step 1) $\sigma$ AESS $\Rightarrow \sigma$ strict. Assume $\sigma$ is an AESS, and choose an arbitrary $i \in \{1, \ldots, n\}$. Let $t_i \in B_i(\sigma)$, and take $\mu = (\sigma \setminus t_i)$. Now, $\pi_i((1-\varepsilon)\sigma + \varepsilon\mu) = \pi_i((1-\varepsilon)\sigma + \varepsilon\mu) \setminus \sigma_i$ because $\mu_i - t_i$ is a best response to $\sigma$. For $j \neq i$, $\sigma_j = \mu_j$, and so clearly $\pi_j((1-\varepsilon)\sigma + \varepsilon\mu) \setminus \sigma_j = \pi_j((1-\varepsilon)\sigma + \varepsilon\mu) \setminus \sigma_j$. Thus, by the definition of AESS, $\mu = \sigma$, and so $\sigma_i(t_i) = 1$ must hold. As $t_i$ was an arbitrary element of $B_i(\sigma)$, $\{t_i\} = B_i(\sigma)$. As $i$ was arbitrary, $\sigma$ is strict. Conversely, if $\sigma$ is strict, then it is trivially an AESS.

(Step 2a) $(\sigma, \sigma)$ an ESS of $(T^2, \Pi) \Rightarrow \sigma$ strict in $(S, \pi)$. Assume $t_i \in B_i(\sigma)$, i.e., $\pi_i(t_i, \sigma_2) = \pi_i(\sigma_1, \sigma_2)$. Then, $2\Pi_1(\sigma, \sigma) = \pi_1(\sigma) + \pi_2(\sigma) = \pi_1(\sigma) + \pi_2(\sigma) + \pi_2(\sigma) = 2\Pi_1((t_1, \sigma_2), \sigma)$. Also, $2\Pi_1(\sigma(t_1, \sigma_2)) = \pi_1(\sigma) + \pi_2(t_1, \sigma_2) = \pi_1(\sigma) + \pi_2(t_1, \sigma_2) = 2\Pi_1((t_1, \sigma_2), (t_1, \sigma_2))$. By the definition of an ESS, this implies $\sigma_i(t_i) = 1$. Arguing as in step 1, and analogously for player 2, $\sigma$ is strict.

(Step 2b) $\sigma$ strict $\Rightarrow (\sigma, \sigma)$ an ESS of $(T^2, \Pi)$. It is sufficient to show that $\sigma$ strict in $(S, \pi) \Rightarrow (\sigma, \sigma)$ strict in $(T^2, \Pi)$. So, let $\mu \in T_1, \mu \neq \sigma$. Then, $2\Pi_1(\sigma, \sigma) - 2\Pi_1(\mu, \sigma) = \pi_1(\sigma) + \pi_2(\sigma) - \pi_1(\sigma_1, \sigma_2) - \pi_2(\sigma_1, \mu_2)$. The first term minus the third is positive because $\{t_1\} = B_1(\sigma)$, where $\sigma(t) = 1$, and similarly for the second term minus the fourth.

Because $(T^2, \Pi)$ is what Selten [12] refers to as a truly asymmetric contest, the second equivalence in the theorem is a special case of Theorem 9.6.2 in [16], originally due to Selten [12].

Evolutionary stability thus is a very strong condition for asymmetric games. However, the same observation that motivated the restriction to equilibrium entrants for symmetric games applies here: if the entrants are
not playing best responses to the population mix, then it seems likely that the entrants' actions will change. If we weaken AESS to reflect this, we obtain:

**Definition 7.** The strategy profile \( \sigma \) is *robust against equilibrium entrants* (REE) in \((S, \pi)\) if \( \exists \epsilon' > 0 \) such that \( \forall \epsilon \in (0, \epsilon') \), and for any \( \mu \),

\[
C(\mu) \subseteq B((1 - \epsilon) \sigma + \epsilon \mu) \Rightarrow \mu = \sigma.
\]

As with the definition of RSEE, an equivalent condition is

1. \( \sigma \) is a Nash equilibrium of \((S, \pi)\), and
2. \( \sigma \) is the only Nash equilibrium of \((B(\sigma), \pi)\).

**Observation 2.** If \( \sigma \) is an AESS for \((S, \pi)\), then it is REE.

Once again, it is a special case of Theorem 5 that if \( \sigma \) is REE for \((S, \pi)\), then \( \sigma \) is a proper equilibrium for \((S, \pi)\).

The remaining implications of robustness to equilibrium entrants discussed in the previous section go through immediately in the asymmetric case.

**VII. Problems with Robustness**

The restriction to equilibrium entrants seems to address the issue of unstable entrants in a simple and fairly natural way, while retaining some of evolutionary stability's attraction. It also is of some help in dealing with asymmetric games. Unfortunately, RSEE and REE share some of evolutionary stability's more troubling problems as well. In particular, strategies that are RSEE or REE fail to exist on many games because of an inability to differentiate strategies that agree in equilibrium. In this section, we will illustrate these problems of REE and RSEE. Because every ESS is also RSEE, these are are a fortiori problems of evolutionary stability as well.

**Definition 8.** Let \( X_i \subseteq S_i \), \( i = 1, \ldots, n \). Then, two strategies \( \mu, v \in \Phi \) agree on \( X = \prod_{i=1}^{n} X_i \) if \( n(s \setminus \mu_i) = n(s \setminus v_i) \ \forall s \in X, \forall i \). A pure strategy \( t \) is *redundant* if it agrees on \( S \) with some \( \sigma \) for which \( t \notin C(\sigma) \).

**Lemma 1.** A necessary condition for \( \sigma \) to be RSEE (REE) in \((S, \pi)\) is that no other element of \( \Phi \) agrees with \( \sigma \) on \( B(\sigma) \).

**Proof.** Let \( \mu \neq \sigma \) agree with \( \sigma \) on \( B(\sigma) \) (note that for a symmetric profile \( \sigma \) in a symmetric game, if there is any such \( \mu \), then there is a symmetric such \( \mu \)). Because \( \sigma \in N(S, \pi) \), \( C(\mu) \subseteq B(\sigma) \). But, then as \( \sigma \in N(B(\sigma), \pi) \), \( \mu \in N(B(\sigma), \pi) \).
Thus, in particular, the existence of a RSEE (REE) strategy profile can depend on the addition or removal of redundant strategies.

**Definition 9.** \((S, \pi)\) is rectangular if \(\forall s, t \in S,\)

\[
(1) \quad [\pi_i(s) = \pi_i(t)] \Rightarrow [\pi_j(s) = \pi_j(t)] \quad \forall i, j = 1, ..., n, \text{ and}
\]

\[
(2) \quad [\pi(s) = \pi(t)] \Rightarrow [\pi(s) = \pi(r)] \quad \text{for any } r \text{ such that } r_i = s_i \text{ or } r_i = t_i \quad \forall i = 1, ..., n.
\]

**Lemma 2.** Let \(\Gamma\) be a generic extensive form, and let \((S, \pi)\) be the normal form of \(\Gamma\) (or any game obtained from the normal form of \(\Gamma\) by the removal or addition of redundant strategies). Then, \((S, \pi)\) is rectangular.

**Proof.** Begin with the normal form of \(\Gamma\). For a generic extensive form game, \(\pi_i(s) = \pi_i(t)\) \(\Rightarrow\) \(s\) and \(t\) generate the same distribution over terminal nodes \(\Rightarrow\) \(\pi_j(s) = \pi_j(t)\). Also, \(\pi(s) = \pi(t)\) \(\Rightarrow\) \(s\) and \(t\) generate the same distribution over terminal nodes \(\Rightarrow\) \(s_i\) and \(t_i\) agree on every relevant information set \(\Rightarrow\) if \(r = (r_1, ..., r_n)\) has \(r_i\) equal to either \(s_i\) or \(t_i\) for each \(i\) then \(r\) generates the same distribution over terminal nodes as do \(s\) and \(t\). It is clear that these properties survive the removal or addition of redundant strategies.

**Lemma 3.** If \((S, \pi)\) is a rectangular symmetric game, then a symmetric pure strategy profile \(s \in S\) is RSEE in \((S, \pi)\) only if it is strict.

**Proof.** Under the assumptions, all strategy profiles in \(B(s)\) yield the same payoff vector. Unless \(B(s)\) is a singleton, Lemma 1 implies that \(s\) is not RSEE.

Thus, for the normal form (or reduced normal form) of a generic extensive form game, a symmetric pure strategy profile \(s\) is RSEE only if it is strict. The profile \((0, 0)\) in the game of Fig. 2 is not a counter-example: this game is not the reduced normal form of any generic extensive form game.

The non-existence of REE strategy profiles is even more pervasive:

**Lemma 4.** A pure strategy profile \(s \in S\) is REE in \((S, \pi)\) only if it is strict.\(^5\)

**Proof.** Let the pure strategy profile \(s \in N(S, \pi)\) be non-strict. Then, without loss of generality, \(B_1(s) \setminus s_1 \neq \emptyset\), so \((B_1(s) \setminus s_1) \times B_2(s) \times \cdots \times B_n(s), \pi)\) has a Nash equilibrium \(\mu\). Now, if \(\mu \in N(B(s), \pi)\), then we are finished, as clearly \(\mu \neq s\). If \(\mu \notin N(B(s), \pi)\) then \(\pi_1(\mu \setminus s_1) > \pi_1(\mu)\). As

\(^5\) This strengthening of Lemma 3 for asymmetric games is due to Michele Piccione.
A related difficulty arises even when one is not restricted to pure strategies. Consider the (reduced normal form) game of Fig. 3. This game corresponds to the extensive form of Figure 4. A reasonable prediction for this game might be \((T, R)\). The set of all Nash equilibria is \(\Theta = \{\sigma | \sigma_1 = (1, 0), \sigma_2 = (1 - r, r), r \geq 1/2\}\). All elements of \(\Theta\) correspond to the same path in the extensive form. Because II's information set is unreached, any action for II is a best response. There is a range of such actions consistent with player I not wishing to deviate. Such an action by II paired with a choice of \(T\) by I is thus an equilibrium entrant, implying that no strategy profile is REE for this game.

This is a general phenomenon. Starting with a Nash equilibrium for an extensive form game \(\Gamma\), any strategy profile that differs from the equilibrium strategy profile only in what it specifies at information sets that are never reached given the equilibrium strategy profile will be a best response. If there is a range of behavior at these unreached information sets consistent with equilibrium, then the associated normal form will fail to have an REE strategy profile (and similarly for RSEE).

In the reduced normal form things are a little more complicated. Essentially, the reduced normal form ignores behavior by a player at an information set that she herself has ruled out. Thus, if the only unreached information sets are ones that the owner of the information set has ruled out reaching, then this problem does not arise. However, if an unreached...
information set is possible given the play of its owner, but not given the play of her opponent, then the difficulty will persist in the reduced normal form.

VIII.i. Equilibrium Evolutionary Stability

The failure of REE (RSEE) strategy profiles to exist for the normal form (or reduced normal form) of many interesting and seemingly "stable" extensive form games motivates us to consider set valued solution concepts.\(^6\) Henceforth, we consider only the case in which the populations from which different players who meet to play the game are drawn are independent. This is for brevity; modifying the discussion to deal with symmetric games with a single population of players is straightforward.\(^7\)

**Definition 10.** A set \( \Theta \subseteq \Phi \) is equilibrium evolutionarily stable (EES) in \((S, \pi)\) if it is minimal with respect to

1. \( \Theta \) is closed,
2. \( \Theta \subseteq N(S, \pi) \),
3. \( \exists \varepsilon > 0 \) such that \( \forall \varepsilon \in (0, \varepsilon') \), \( \forall \sigma \in \Theta \), and \( \forall \mu \), \( C(\mu) \subseteq B((1 - \varepsilon) \sigma + \varepsilon \mu) \Rightarrow (1 - \varepsilon) \sigma + \varepsilon \mu \in \Theta \).

So, an EES set is a closed set of Nash equilibria that cannot be left by equilibrium entrants making up less than \( \varepsilon' \) of the population. It is clear that a set \( \Theta \) with a single element \( \sigma \) will be EES if and only if \( \sigma \) is REE.

Condition (1) is largely a convenience: with modifications, the analysis that follows would go through, but with complications to theorem statements and proofs. We insist on Condition (2) although it plays no crucial role in much of the analysis that follows and is the reason that EES sets need not exist on every game. The importance of Condition (2) is to the interpretation of an EES set: Condition (3) represents a way for entrants to choose strategies that is intermediate between the total absence of rationality imposed on entrants by evolutionary stability and a fully rational choice based on a complete estimation of the future evolution of the population strategy profile. If a set satisfying Condition (3) also satisfies (2), and if all (present and future) entering players choose in this way, then entering players will never have cause to regret their decision.

\(^6\) An alternate approach to extending evolutionary stability to extensive form games is that suggested by Selten [13]. The approach there is to look at evolutionarily stable strategies of perturbed games in which players are constrained to play all strategies with positive probability. An example of a game on which the two concepts differ is the beer–quiche game of Cho and Kreps [2]. See Fig. 6, below, and the accompanying discussion.

\(^7\) Gilboa and Matsui [3] allow for intermediate cases in which, for example, a game is played by three players, two of whom are drawn from one population, and one from another. The analysis is easily extended to this case.
i.e., choosing in this myopic fashion is an optimal strategy. Thus, there is at least an internal consistency to assuming entering players choose according to (3). If, on the other hand, players who enter according to (3) can take the population out of the set of Nash equilibria, then there is an incentive for players to attempt to make decisions in a less myopic way. I thus find it quite implausible that a population that is playing an element of a set satisfying (3) but not (2) would chose to continue making entry decisions according to (3). Put differently, in the space of decision mechanisms, choosing entering strategies according to (3) is a Nash equilibrium if (2) is also satisfied, but not otherwise. Ideally, of course, one would like a complete theory of how the population strategy evolves, even when there is no set satisfying both (3) and (2). However, I find a theory that predicts that the population will stay in a set solely because it satisfies (3) unsatisfactory.

We were able to reformulate the definitions of REE and RSEE in terms of a condition involving $N(B(\sigma), \pi)$. For EES, one might analogously replace Condition (3) with

$$(3') \forall \sigma \in \Theta, \mu \in N(B(\sigma), \pi) \Rightarrow \exists \varepsilon > 0 \text{ such that } (1 - \varepsilon) \sigma + \varepsilon \mu \in \Theta \forall \varepsilon \in (0, \varepsilon').$$

It is straightforward to show that using (3') is no stronger than using (3): Let $\Theta$ be an EES set. As was argued previously, for each $\mu$, there exists $\delta$ such that for $\varepsilon < \delta$, $C(\mu) \subseteq B((1 - \varepsilon) \sigma + \varepsilon \mu) \Leftrightarrow C(\mu) \subseteq B_{W}(\mu, \varepsilon) \Leftrightarrow \mu \in N(B(\sigma), \pi)$. Thus, if $\sigma \in \Theta$, then taking $\varepsilon \leq \min(\varepsilon', \delta)$, $\mu \in N(B(\sigma), \pi) \Rightarrow (1 - \varepsilon) \sigma + \varepsilon \mu \in \Theta$.

It is unclear whether (3') is as strong as (3). While condition (3') implies that for each $\sigma$ there is an $\varepsilon'$ such that $\sigma$ is robust to mutations of size less than $\varepsilon'$, it is not obvious that $\varepsilon'$ can be chosen independently of $\sigma$. My conjecture is that it can, which would be useful as Condition (3') is easier to check than (3).

The equivalence of (3) and (3') also would be of interest because Matsui [7] shows that sets closed under the modified Gilboa–Matsui dynamic are precisely those minimal with respect to Conditions (1) and (3').

The example of Fig. 3 illustrates the operation of this solution concept (EES sets are illustrated for a number of more elaborate games in Section VIII.iii). For this game, the unique EES set is the set of all Nash equilibria, i.e., the set $\Theta = \{(\sigma_1, \sigma_2) | \sigma_1 = (1, 0), \sigma_2 = (1 - r, r), \text{ where } r \geq \frac{1}{2}\}$. Conditions 1 and 2 are obvious. That $\Theta$ is minimal is clear because for all $\sigma$ such that $\sigma(R) > \frac{1}{2}$, $B(\sigma) = T \times S_2$, so a Nash equilibrium of $(B(\sigma), \pi)$ is $(T, L)$. Any closed set of Nash equilibria for which the minimal value of $\sigma(R)$ is greater than $\frac{1}{2}$ thus fails (3') and so also fails (3) (this is a special case of Theorem 3 below). Finally, note that $\forall \sigma \in \Theta, \forall \varepsilon \in [0, 1]$, and $\forall \mu \in \Phi, C(\mu) \subseteq B((1 - \varepsilon) \sigma + \varepsilon \mu) \Rightarrow \mu \in \Theta$. If player II's payoffs to $(B, L)$ and
(B, R) were reversed, then \( \forall \varepsilon > 0, \mu = (B, L) \) would fail Condition (3) (let \( \sigma \in \Theta \) satisfy \( (1 - \varepsilon) \sigma_2(R) \leq \frac{1}{2} \)).

EES sets are not guaranteed to exist, as is illustrated by the game of Fig. 5.

This game has 3 Nash equilibria: (T, L) with payoffs (2, 2), (M, C) with payoffs (3, 3), and a mixed strategy Nash equilibrium \((\left( \frac{13}{28}, \frac{15}{28}, \frac{6}{28} \right), \left( \frac{13}{28}, \frac{15}{28}, \frac{6}{28} \right))\) with payoffs \(\left( \frac{27}{14}, \frac{27}{14} \right)\). Any EES set must be a made up of one or more of these three Nash equilibria. However, it is easily verified that for \( \varepsilon \) small, \((M, C) \subseteq B((1 - \varepsilon)(T, L) + \varepsilon(M, C))\). Since for \( \varepsilon \in (0, 1) \) \( (1 - \varepsilon)(T, L) + \varepsilon(M, C) \) is not a Nash equilibrium, \((T, L)\) cannot be a part of any EES set. Similarly, \((B, R)\) is an equilibrium entrant for a population starting at \((M, C)\) and \((T, L)\) an equilibrium entrant for a population starting at the mixed equilibrium, and so these equilibria cannot be part of an EES set either. But, then there can be no EES set, as we have ruled out all the Nash equilibria. As the entrants and the equilibria mentioned are all symmetric, this example also shows non-existence for the symmetric version of EES.

We now turn to the structure of an EES set.

**Theorem 3.** Let \( \Theta \) be an EES set. Then, \( \Theta \) is equal to a maximal connected set of Nash equilibria.

**Proof.** Let \( \Theta \) be an EES. First note that for \( \varepsilon > 0, (1 - \varepsilon) \sigma + \varepsilon \mu \in N(S, \pi) \Rightarrow C(\mu) \subseteq B((1 - \varepsilon) \sigma + \varepsilon \mu) \). So, if \( \sigma \in \Theta \), and \( (1 - \varepsilon) \sigma + \varepsilon \mu \in N(S, \pi) \) for \( 0 < \varepsilon < \varepsilon' \), \( \mu \in \Phi \), then \( (1 - \varepsilon) \sigma + \varepsilon \mu \in \Theta \). Next, note that for each \( \sigma \in \Theta \), there exists \( \delta > 0 \) such that when \( D(\sigma, \eta) < \delta \), there exists \( \mu \in \Phi \) and \( \varepsilon < \varepsilon' \) such that \( \eta = (1 - \varepsilon) \sigma + \varepsilon \mu \). Combining these two facts, we have that there is a neighborhood around each \( \sigma \in \Theta \) such that Nash equilibria in that neighborhood are also in \( \Theta \). Taking a union of such neighborhoods over \( \Theta \), we construct an open set \( U \) with \( \Theta \subseteq U \) and \( N(S, \pi) \cap U \subseteq \Theta \). Now, since \( \Theta \) is closed \( V \equiv \Phi \setminus \Theta \) is open. \( V \cap U = U \setminus \Theta \), and so \( N(S, \pi) \cap V \cap U = \emptyset \). Thus, as \( N(S, \pi) \subseteq U \cup V = \Phi \), any set of Nash equilibria having elements both in \( \Theta \) and not in \( \Theta \) has \( U, V \) as a separation (see for example [11, p.188]). Hence, any connected set of Nash equilibria having an element in \( \Theta \) is contained in \( \Theta \).

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>C</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>2,2</td>
<td>0,2</td>
<td>5,0</td>
</tr>
<tr>
<td>I</td>
<td>2,0</td>
<td>3,3</td>
<td>0,3</td>
</tr>
<tr>
<td>B</td>
<td>0,5</td>
<td>3,0</td>
<td>4,4</td>
</tr>
</tbody>
</table>

**Figure 5**
Also note that for $\Theta$ to contain more than one component of Nash equilibria would violate minimality: for $\varepsilon$ sufficiently small, $\Theta$ will satisfy Condition (3) of Definition 10 if and only if each of its maximal connected subsets does.

An implication of this result is that generically, all elements of an EES set will generate the same outcome in any associated extensive form, and the same payoffs [6, Theorem 2].

Part of an EES set always survives the iterative removal of weakly dominated strategies in any order:

**Definition 11.** For a subset $X = \prod_{t=1}^{n} X_t$ of $S$, define the function $X(\cdot): \Phi \to [0, 1]$ by $X(\eta) = \sum_{x \in X} \eta(x)$. Thus $X(\eta) = 1 \iff C(\eta) \subseteq X$.

**Theorem 4.** Let $\Theta$ be an EES set in the game $(S, \pi)$. Let $W = \bigcap_{k=0}^{K} W_k$, where $W_0 = S$, and $\forall k = 1, \ldots, K$, $W_{k+1}$ is obtained from $W_k$ by the removal of some number (possibly zero) of pure strategies that are weakly dominated in $(W_k, \pi)$. Then, there exists $\sigma \in \Theta$ such that $W(\sigma) = 1$.

**Proof.** By induction on $k$. Define $\Theta_k \equiv \{ \sigma \in \Theta \mid W_k(\sigma) = 1 \}$. Then, $\Theta_0 \neq \emptyset$ because $S(\sigma) = 1 \quad \forall \sigma \in \Theta$. So, assume $\Theta_k \neq \emptyset$. Let $\eta = \arg\max_{\sigma \in \Theta_k} W_{k+1}(\sigma)$, and assume $W_{k+1}(\eta) < 1$. Now, $C(\eta) \subseteq W_k$, so $s \in B(\eta) \Rightarrow t \in B(\eta)$ for any $t$ that weakly dominates $s$ on $W_k$. Thus, $B(\eta) \cap W_{k+1} \neq \emptyset$. Let $\gamma \in N(B(\eta) \cap W_{k+1}, \pi)$. Then, $\gamma \in N(B(\eta), \pi)$. So, by the definition of an EES set, $\exists \varepsilon > 0$ such that $\xi \equiv (1 - \varepsilon) \eta + \varepsilon \gamma \in \Theta$. But, as $W_{k+1} \subseteq W_k$, $\xi \in \Theta_k$. This is a contradiction as $W_{k+1}(\xi) > W_{k+1}(\eta)$.

EES sets are robust to the addition or removal of redundant strategies:

**Lemma 5.** Let $\Theta$ be an EES set in $(S, \pi)$. Then, $[\sigma \in \Theta, \gamma$ and $\sigma$ agree $] \Rightarrow \gamma \in \Theta$.

**Proof.** $\sigma$ is a Nash equilibrium of $(S, \pi)$. Because $\sigma$ and $\gamma$ agree, $(1 - \lambda) \sigma + \lambda \gamma$ is a Nash equilibrium for $\lambda \in [0, 1]$. Thus, $\sigma$ and $\gamma$ are part of the same connected set of Nash equilibria.

**Lemma 6.** Let $(S, \pi)$ be obtained from $(T, \pi')$ by the addition or removal of a redundant strategy. Let $\Theta^\pi$ be an EES set for $(S, \pi)$, and let $\Theta^T$ be the set of all mixtures over $T$ that agree with an element of $\Theta^\pi$. Then, $\Theta^T$ is an EES set of $(T, \pi')$.

**Proof.** Conditions (1) and (2) are obvious. Let $\sigma' \in \Theta^T$, and let $C(\mu') \subseteq B((1 - \varepsilon) \sigma' + \varepsilon \mu')$, where $\varepsilon < \varepsilon'$ from Condition (3). Then, $\exists \sigma \in \Theta^\pi$ and $\mu$ such that $\sigma$ agrees with $\sigma'$ and $\mu$ agrees with $\mu'$. Thus, $C(\mu) \subseteq B((1 - \varepsilon) \sigma + \varepsilon \mu)$ and so $(1 - \varepsilon) \sigma + \varepsilon \mu \in \Theta^\pi$. As $(1 - \varepsilon) \sigma + \varepsilon \mu$ and $(1 - \varepsilon) \sigma' + \varepsilon \mu'$ agree, $(1 - \varepsilon) \sigma' + \varepsilon \mu' \in \Theta^T$. \[\square\]
Thus, the set of payoffs sustainable by EES sets does not depend on the addition or removal of redundant strategies.

VIII.ii. EES Sets and Properness

A surprising and powerful property of both evolutionary stability and robustness to equilibrium entrants (symmetric and asymmetric forms) is that they imply properness. In this section, we prove that every EES set with a singleton element, and every EES set in the normal form of a generic two person extensive form game, contains a proper element. The two major tasks in this section are to show that for the normal forms of extensive two person games (i) maximal connected sets of Nash equilibria are convex, and (ii) no entrant that moves the population less than a certain Euclidean distance can take the population out of the EES set, regardless of how large a fraction of the population that entrant makes up. The actual proof that EES sets having these two properties contain a proper element is quite simple, and is in fact a special case of the central result of [15], where it is established that these properties are enough to imply the existence of stable subsets, under both the Kohlberg and Mertens [5] and Hillas [4] formulations. A proof of the special case is provided both for completeness and to provide a proof that does not rely on the considerable machinery of stability.

We begin with a simple lemma:

**Lemma 7.** Let \((S, \pi)\) be a two person game, and assume \(\mu, \nu \in \Phi\) satisfy \(B(\mu) \cap B(\nu) \neq \emptyset\). Then, \(\forall \lambda \in (0, 1), B\left((1 - \lambda)\mu + \lambda\nu\right) = B(\mu) \cap B(\nu)\).

**Proof.** \((\supseteq)\) As \(\pi_1(s_1, (1 - \lambda)\mu_2 + \lambda\nu_2) = (1 - \lambda)\pi_1(s_1, \mu_2) + \lambda\pi_1(s_1, \nu_2)\) \(\forall s_1 \in S_1, \forall \lambda \in [0, 1]\) (and similarly for player 2) it is clear that \(r \in B(\mu) \cap B(\nu) \Rightarrow r \in B\left((1 - \lambda)\mu + \lambda\nu\right)\).

\((\subseteq)\) Assume \(r \in B\left((1 - \lambda)\mu + \lambda\nu\right)\) where \(\lambda \in (0, 1)\). Let \(t \in B(\mu) \cap B(\nu)\). Then, we have

\[
\pi_1(t_1, (1 - \lambda)\mu_2 + \lambda\nu_2) = (1 - \lambda)\pi_1(t_1, \mu_2) + \lambda\pi_1(t_1, \nu_2) \\
(1 - \lambda)\pi_1(t_1, \mu_2) + \lambda\pi_1(t_1, \nu_2) \\
= \pi_1(t_1, (1 - \lambda)\mu_2 + \lambda\nu_2) \\
\leq \pi_1(r_1, (1 - \lambda)\mu_2 + \lambda\nu_2).
\]

The reason why there is anything to show here is that an entrant of size \(\varepsilon\) can only move a population a fraction \(\varepsilon\) of the distance from its beginning position toward any particular boundary. Thus, as one approaches any particular boundary, a movement of any particular distance toward that boundary will require entrants to make up a larger and larger fraction of the population. Showing that EES sets in two person games which come from generic extensive forms have this property is equivalent to showing that they are uniformly robust to equilibrium entrants, defined in the companion paper [15].
The weak inequalities are thus equalities. As \( \lambda \neq 0 \) or 1, the second equality implies \( r_1 \in B_1(\mu) \cap B_1(v) \). Arguing analogously for player 2, \( r \in B(\mu) \cap B(v) \).

For games with more than two players, it is no longer the case that
\[
\pi_1(r_1, (1 - \lambda) \mu_{-1} + \lambda v_{-1}) = (1 - \lambda) \pi_1(r_1, \mu_{-1}) + \lambda \pi_1(r_1, v_{-1}).
\]
Thus, this and dependent results depend critically on the restriction to two player games. We have:

**Lemma 8.** If \((S, \pi)\) is the normal form of a two player generic extensive form game (or is obtained from such a normal form by the addition or removal of redundant strategies), then any maximal connected set of Nash equilibria (and thus any EES set) of \((S, \pi)\) is convex.

**Proof.** If a particular strategy \( \sigma \) satisfies \( C(\sigma) \subseteq B(\sigma) \), then any strategy \( \gamma \) that differs from \( \sigma \) only at information sets not reached by \( \sigma \) will satisfy \( C(\gamma) \subseteq B(\sigma) \). So let \( \sigma, \mu \) belong to a connected set of Nash equilibria of a generic extensive form. Then, \( \sigma \) and \( \mu \) generate the same distribution across terminal nodes (see Kreps and Wilson [6, Theorem 2]) and so reach the same information sets and specify the same actions at reached information sets. Thus any convex combination \( \eta \) of \( \sigma \) and \( \mu \) will differ from \( \sigma \) and \( \mu \) only at unreached information sets, and so \( C(\eta) \subseteq B(\sigma) \cap B(\mu) \). Because this is a two player game, \( B(\sigma) \cap B(\mu) = B(\eta) \) by Lemma 7, and we are done.

Let \( \Lambda = \{v \in \Phi \mid \exists \sigma \in \Theta \text{ such that } B(v) \subseteq B(\sigma)\} \). That is, \( \Lambda \) is the set of all strategy profiles that have best responses a subset of the best responses to some point in \( \Theta \). We show that no equilibrium entrant of any size can move the population from a convex EES set in a two person game (and therefore from an EES set for the normal form of a generic two person extensive form game) to a point in \( \Lambda \setminus \Theta \).

**Lemma 9.** Let \( \Theta \) be a convex EES set of a two person game. Let \( \xi \in \Lambda \) and let \( \gamma \in \Phi, \eta \in \Theta, \) and \( \alpha \in [0, 1] \) be such that \( \xi = (1 - \alpha) \eta + \alpha \gamma \). Then,
\[
[C(\gamma) \subseteq B(\xi)] \Rightarrow \xi \in \Theta.
\]

**Proof.** Assume the lemma is false, i.e., assume there exists \( \xi \in \Lambda \setminus \Theta, \gamma \in \Phi, \eta \in \Theta \), and \( \alpha \in [0, 1] \) such that \( \xi = (1 - \alpha) \eta + \alpha \gamma \) and \( C(\gamma) \subseteq B(\xi) \). By definition of \( \Lambda \) there exists \( \sigma \in \Theta \) such that \( B(\xi) \subseteq B(\sigma) \). Because \( \Theta \) is closed, \( \sigma \) can be chosen such that \( (1 - \lambda) \sigma + \lambda \xi \notin \Theta \) for any \( \lambda \in [0, 1] \). (If \( \sigma \) does not have this property, then choose the point of the form \( (1 - \lambda) \sigma + \lambda \xi \notin \Theta \), \( \lambda \in [0, 1] \) that is closest to \( \xi \) while still in \( \Theta \). By Lemma 7, this point has the required best responses.)
Let $\lambda \in (0, 1)$, and consider the point

$$(1 - \lambda) \sigma + \lambda \xi = (1 - \lambda) \sigma + \lambda [(1 - \alpha) \eta + \alpha \gamma].$$

where $\phi = ((1 - \lambda)/(1 - \lambda \alpha)) \sigma + (\lambda(1 - \alpha)/(1 - \lambda \alpha)) \eta$. $\phi$ is a convex combination of elements of $\Theta$ and therefore $\phi \in \Theta$. By Lemma 7, $\forall \lambda \in (0, 1)$, $B(\xi) = B((1 - \lambda \alpha) \phi + \lambda \alpha \gamma)$. As $C(\gamma) \subseteq B(\xi)$, we have $C(\gamma) \subseteq B((1 - \lambda \alpha) \phi + \lambda \alpha \gamma)$ for $\phi$ an element of $\Theta$. As $\Theta$ is an EES set, this implies $(1 - \lambda) \sigma + \lambda \xi \in \Theta$ for $\lambda < \varepsilon / \alpha$. But, for $\lambda > 0$, this contradicts the choice of $\sigma$, yielding the result. 

Thus, the maximum $\varepsilon$ associated with an EES will, for many examples, be quite large. For the games of Figs. 2, 3, and 8 (below), $\Lambda = \emptyset$, and thus the "entrants" can make up an arbitrary proportion of the population.

A final lemma gives the distance result:

**Lemma 10.** There exists $\delta > 0$ such that for all $\sigma \in \Theta$, $\eta \in \Phi$, and $D(\sigma, \eta) \leq \delta \Rightarrow \eta \in A$.

**Proof.** Define $T = \{ \Psi \subseteq S : \forall \sigma \in \Theta, \Psi \not\subseteq B(\sigma) \}$. For each $\Psi \in T$, $V(\Psi) \equiv \{ v : \Psi \subseteq B(v) \}$ is either empty or a closed subset of $\Phi$ that is disjoint from $\Theta$ and thus has $D(V(\Psi), \Theta) > 0$. As $|T| \leq 2^{|S|}$, and is thus finite, $\min_{v \in T, v(\Psi) \neq \emptyset} D(V(\Psi), \Theta)$ is defined and positive, establishing the result.

We can now prove the following:

**Theorem 5.** If $\Theta$ either contains a single element or is an EES set of the normal form $(S, \pi)$ of a generic two person extensive form game, then $\Theta$ contains a proper equilibrium of $(S, \pi)$.

**Remarks.** The modification of this proof to the symmetric version of EES is trivial. As AESS $\Rightarrow$ REE $\Rightarrow$ singleton EES, and similarly for the symmetric case, this also implies that strategy profiles satisfying these various solution concepts are proper. The companion paper [15] generalizes this result in a number of ways.

**Proof.** Under the conditions of the theorem, $\Theta$ is convex, either because it is a singleton, or by Lemma 8. Thus by Lemma 10, we can choose $\varepsilon > 0$ such that any $\eta$ within $\varepsilon$ of $\Theta$ has $\eta \in A$. Define $Y = \{ v : D(\sigma, \Theta) \leq \varepsilon \}$. Since $\Theta$ is ESS by assumption, Lemma 9 then implies that no equilibrium entrant can move the population from within $\Theta$ to $Y \setminus \Theta$.
We work with the mapping $P_\delta: \Theta \to 2^\Theta$ defined for $0 < \delta < 1$ by

$$P^\delta_\delta(\sigma) = \left\{ \eta \mid \begin{align*}
(1) & \eta(s_i) \geq \frac{\delta^{1\forall i}}{|S_i|} \forall i = 1, \ldots, n \\
(2) & \pi^*_i(\sigma \setminus s_i) > \pi_i^*(\sigma \setminus t_i) \Rightarrow \eta_i(t_i) \leq \eta_i(s_i) \forall i = 1, \ldots, n.
\end{align*} \right\}$$

Following Myerson [10], for each $\delta$, $P_\delta$ is an upper hemicontinuous, convex valued, non-empty correspondence on $\Theta$. Thus, $P_\delta$ has a fixed point for each $\delta$. Taking a convergent subsequence of such fixed points as $\delta \downarrow 0$ yields a proper equilibrium. We desire such a subsequence that converges to an element of $\Theta$.

For any point $\sigma$, let $V(\sigma) = \arg\min_{v \in \Theta} D(\sigma, v)$. Because $\Theta$ is closed and convex, and $D$ strictly convex, $V$ is uniquely defined and continuous. Let

$$L(\sigma) = \{ \mu | \mu \in Y, \exists \sigma \in [0, 1] \text{ such that } \mu = (1 - \alpha) V(\sigma) + \alpha \sigma \}. $$

Because $L(\sigma)$ is continuous and convex valued as a correspondence, and $D$ is convex, $A(\sigma) = \arg\min_{\mu \in L} D(\sigma, \mu)$ is upper hemicontinuous. Again using the strict convexity of $D$, $A$ is singleton valued and thus continuous. By construction, $A$ is onto $Y$, $A|_Y$ is the identity map, and no point outside $Y$ gets mapped to the interior of $Y$.

Define $R_\delta(\sigma) = P_\delta(A(\sigma))$. Because $A$ is continuous, $R_\delta$ inherits the relevant properties of $P_\delta$, and thus $R_\delta$ has a fixed point for each $\delta$. For a sequence $\delta \downarrow 0$, choose a fixed point $\sigma^\delta$ for each $\delta$. Choose a convergent subsequence, and call its limit $\sigma$. Now, we have $\pi_i(A(\sigma^\delta) \setminus s_i) < \pi_i(A(\sigma^\delta) \setminus t_i) \Rightarrow \sigma^\delta_i(s_i) < \delta \sigma^\delta_i(t_i) \forall s_i, t_i \in S_i, \forall \delta$, and similarly for player 2. Taking limits, we have $C(\sigma) \in B(A(\sigma))$. As $A(\sigma) = (1 - \alpha) V(\sigma) + \alpha \sigma$ for some $\alpha \in [0, 1]$, as $V(\sigma) \in \Theta$, and as $A(\sigma) \in A$, Lemma 9 implies that $A(\sigma) \in \Theta$ and thus $\sigma = A(\sigma)$. For $\delta$ small, $\sigma^\delta \in Y$ and so $A(\sigma^\delta) = \sigma^\delta$. $R_\delta$ and $P_\delta$ then agree for $\delta$ small, and $\sigma$ is a proper equilibrium of $(S, \pi)$. 

**VIII.iii. EES, the NWBR Property, and Forward Induction**

Two related properties of a solution concept are the never a weak best response (NWBR) property [5] and forward induction [5, 17]. We show that EES sets satisfy the NWBR property and capture some but not all of the implications of forward induction. This section serves the dual role of illustrating the operation of the EES concept for some more elaborate games.

**Definition 12.** A set valued solution concept satisfies the never a weak best response (NWBR) property (Kohlberg and Mertens [5]) if every equilibrium set $\Theta$ contains an equilibrium set of a game obtained by the removal of a strategy that is not a best response to any element of $\Theta$. 

THEOREM 6. Equilibrium evolutionary stability satisfies the NWBR property.

Proof. If \( s \) is never a weak best response, then \( s \notin B(\eta) \) for any \( \eta \) sufficiently close to \( \Theta \). Thus, for \( \varepsilon \) sufficiently small, \( \sigma \in \Theta \), and \( \mu \in \Phi \),

\[
B_{S,N}(1 - \varepsilon) \sigma + \varepsilon \mu = B((1 - \varepsilon) \sigma + \varepsilon \mu),
\]

and so \( \mu \) an equilibrium entrant in \((S, \pi) \Leftrightarrow \mu \) an equilibrium entrant in \((S \setminus s, \pi) \).

It is clear from the proof that EES sets are also unaffected by the addition of strategies that are never a weak best response.

This result can be illustrated by the Beer-Quiche game (Figure I of Cho and Kreps [2]) of Fig. 6.

This game has two sets of Nash (and sequential) equilibria. In the first set, both types of player I drink beer, and no duelling then takes place. In the event that quiche is eaten, duelling is chosen with a probability of at least \( \frac{1}{2} \). In the second set, both types of player I eat quiche, quiche eaters are not made to duel, and beer drinking elicits duelling with probability at least \( \frac{1}{2} \).

The second type of sequential equilibrium fails the NWBR criterion. The equilibrium payoff to a wimpy player I (3) is greater than the maximum payoff a wimpy player I could hope for by drinking beer (2). Thus, any strategy that specifies drinking beer when a wimp is NWBR. Once such strategies are pruned from the game, player II's must interpret beer as indicating surliness, and thus must choose not to duel beer drinkers. Surly player I's then have a positive incentive to deviate.

The set of quiche eating equilibria also fails to be EES. The beer information set is unreached in these equilibria, and thus any response at this information set is a best response. The player II action following beer can thus "drift" in the direction of not duelling. Now, nowhere on (or near) this set of Nash equilibria is it a best response to chose beer if a wimp. So, as the probability of duelling a beer drinker falls below \( \frac{1}{2} \), only surly player I's would choose to deviate. Thus, if one arrives at a situation in which player

![Figure 6](image-url)
II's duel less than $\frac{1}{3}$ the time, and player I's deviate optimally, then those player II's who moved toward not duelling will be happy with their decision not to duel, and the drift toward not duelling will be validated.

The set of beer drinking equilibria is EES. Here "drift" takes place at the quiche information set. If the probability of duelling a quiche eater fell below $\frac{1}{3}$, only wimps would optimally chose to deviate and eat quiche. Player II's would then have cause to regret the choices that brought the probability of duelling a quiche eater below $\frac{1}{3}$.

This game also provides an example in which the outcomes corresponding to EES sets in an underlying extensive form differ from those allowed by Selten's limit ESS [13]. Both equilibrium outcomes can be supported as limit ESSs in this game.

In games where forward induction reasoning has implications stronger than NWRR, EES may or may not capture the stronger restriction. For the extensive form game of Fig. 7 (a small modification of [17] Fig. 4), consider the (sequential) equilibrium in which I plays $T$, and the equilibrium of the subgame is $(\frac{3}{2}, \frac{1}{2}) \times (\frac{1}{2}, 0, \frac{1}{2})$.

This equilibrium fails forward induction. There is a unique equilibrium $(M, L)$ yielding player I a payoff greater than 2 in the subgame. According to the forward induction logic, player II should view being reached in the

<table>
<thead>
<tr>
<th>II</th>
<th>L</th>
<th>C</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>2,2</td>
<td>2,2</td>
<td>2,2</td>
</tr>
<tr>
<td>I M</td>
<td>3,3</td>
<td>1,0</td>
<td>0,2</td>
</tr>
<tr>
<td>B</td>
<td>0,0</td>
<td>0,3</td>
<td>3,2</td>
</tr>
</tbody>
</table>

**Figure 8**
subgame as conclusive evidence that the equilibrium being played in the subgame is actually \((M, L)\).

The normal form of this game is given by Figure 8.

The equilibrium discussed above corresponds to the (proper) equilibrium \(((1, 0, 0), (\frac{1}{2}, 0, \frac{1}{2}))\) of Fig. 8, which is an element of the set \(\Theta\) of all Nash equilibria in which I plays \(T\). \(\Theta\) has the form \(\Theta = T \times \Theta_2\), where \(\Theta_2\) is exhibited in Fig. 9.

Note that there are no weak dominance relations in this game and that every strategy is a weak best response to some element of \(\Theta\). Nonetheless, \(\Theta\) fails to be EES: at \(\sigma = ((1, 0, 0), (\frac{1}{2}, 0, \frac{1}{2})) \in \Theta\), \(B(\sigma) = \{T, M\} \times S_2\). A Nash equilibrium of \((B(\sigma), \pi)\) is \((M, L)\). In this game, EES captures the forward induction restriction. Forward induction and EES are both satisfied for (the strict equilibrium) \((M, L)\).

To see that an EES set need not always contain an element consistent with forward induction, consider the game of Fig. 10.
This game corresponds to an extensive form game in which player I chooses between $(6, 6)$ and the $3 \times 3$ simultaneous move subgame corresponding to $\{1, 2, 3\} \times S_2$. This game has two EES sets. The first is the singleton $\left(\left(\frac{1}{2}, \frac{1}{2}, 0, 0\right), (0, \frac{1}{2}, \frac{1}{2})\right)$ yielding payoffs of $(7.5, 7.5)$. The second is the set of all Nash equilibria in which player I plays only his 4th strategy. This set has the form $4 \times \Theta_2$ where $\Theta_2$ is illustrated in Fig. 11.

The first EES set described satisfies forward induction. The second does not. The unique equilibrium of the subgame corresponding to $\{1, 2, 3\} \times S_2$ that gives payoffs to I of greater than 6 is $\left(\left(\frac{1}{2}, \frac{1}{2}, 0\right) \times (0, \frac{1}{2}, \frac{1}{2})\right)$. Forward induction then says that this must be the equilibrium played in the subgame, upsetting player I's choice not to enter the subgame. However, because strategies 1 and 2 are never simultaneously best responses to any element of $\Theta_2$, this equilibrium does not come into play in determining the equilibrium evolutionary stability of $\Theta$. It should be noted that the subgame has a very non-generic structure, having in particular an even number of equilibria (2). Whether EES sets always capture the forward induction restriction for generic extensive form games is an open question.

REFERENCES