



ELSEVIER

Economics Letters 57 (1997) 45–50

**economics
letters**

Some relationships between evolutionary stability criteria in games

Martin Dufwenberg^{a,b,*}

^aCenter for Economic Research, Tilburg University, Tilburg, Netherlands

^bDepartment of Economics, Uppsala University, Sweden

Received 7 January 1997; received in revised form 13 May 1997; accepted 2 June 1997

Abstract

The relations between several evolutionary stability criteria are studied. © 1997 Elsevier Science S.A.

Keywords: Evolutionary stability

JEL classification: C72

1. Introduction

I study the relations between three old and two new criteria of evolutionary stability in games: evolutionary stability (Maynard-Smith and Price, 1973), neutral stability (Maynard-Smith, 1982), robustness against equilibrium entrants (Swinkels, 1992), robustness against admissible entrants, and robustness against iteratively admissible entrants. Combining new and old findings, the following picture emerges: Neutral stability is neither weaker nor stronger than robustness against equilibrium entrants. Each of these two criteria is neither weaker nor stronger than any of the two new criteria. Evolutionary stability is stronger than the other four criteria taken together. In some games no strategy satisfies any of the criteria.

Section 2 contains results on the old criteria. Section 3 introduces, motivates, and states results involving the two new criteria. Section 4 summarizes most results in a table. Throughout, the analysis deals with mixed extension of finite symmetric 2-player normal form games (see Weibull, 1995 for a detailed presentation of such structures). The following notation is used: In any given game, $A \neq \emptyset$ denotes either of the players' set of (mixed) strategies (infinite-action unless singleton), and $u: A \times A \rightarrow \mathfrak{R}$ denotes the associated payoff function. With $x, y \in A$, $u(x, y)$ is the payoff to strategy x when paired against strategy y . The notation $\varepsilon \cdot y + (1 - \varepsilon) \cdot x$ is used to denote the strategy which puts probability weight ε on y , etc.

*Tel.: +46 18 182309; fax: +46 18 181478; e-mail: martin.dufwenberg@nek.uu.se

2. Results on established criteria

The following well-known definitions specify conditions under which a strategy may be deemed more or less impregnable to evolutionary force:

Definition 1. (Maynard-Smith and Price, 1973): $x \in A$ is an evolutionarily stable strategy (ESS) if $\forall y \in A, y \neq x$ it holds that (i) $u(x,x) \geq u(y,x)$ and (ii) $u(x,x) = u(y,x) \Rightarrow u(x,y) > u(y,y)$.

Definition 2. (Maynard-Smith, 1982): $x \in A$ is a neutrally stable strategy (NSS) if $\forall y \in A, y \neq x$ it holds that (i) $u(x,x) \geq u(y,x)$ and (ii) $u(x,x) = u(y,x) \Rightarrow u(x,y) \geq u(y,y)$.

Definition 3. (Swinkels, 1992): $x \in A$ is robust against equilibrium entrants (REE) if $\exists \varepsilon^* \in (0,1)$ such that $\forall y \in A, y \neq x$ and $\forall \varepsilon \in (0, \varepsilon^*)$ it holds that $y \notin \operatorname{argmax}_{z \in A} u(z, \varepsilon \cdot y + (1 - \varepsilon) \cdot x)$.

It is immediate from the definitions that if $x \in A$ is ESS then x is NSS. Swinkels (1992) shows that if $x \in A$ is ESS then x is REE. Weibull (1995), Section 2.3.2, points out that in a game such that $u(a,b) = u(c,d) \forall a,b,c,d \in A$ (with A not singleton) any strategy is NSS but no strategy is REE, so neutral stability does not imply robustness against equilibrium entrants. The game G_1 shows that no implication runs the other way:

	a	b	c	
a	1,1	0,1	1,1	
b	1,0	1,1	0,2	
c	1,1	2,0	0,0	G_1

In G_1 , the strategy a is not NSS (b can ‘invade’ in the sense that condition (ii) of Definition 2 is violated). However, a is REE since $\forall y \in A, y \neq a$ and $\forall \varepsilon > 0$ it holds that $y \notin \operatorname{argmax}_{z \in A} u(z, \varepsilon \cdot y + (1 - \varepsilon) \cdot a)$, as is easily verified.

A strategy may be both NSS and REE without being ESS. This is an immediate consequence of the last result of the next section (base on G_6).

3. Results involving two new criteria

In order to motivate the two new stability criteria that will be introduced in this section, it is convenient to first consider a reformulation of the REE concept:

Definition 3’: $x \in A$ is robust against equilibrium entrants (REE) if $\exists \varepsilon^* \in (0,1)$ such that $\forall y \in A, y \neq x$ it holds that (i) $u(x,x) \geq u(y,x)$, and (ii) $u(x,x) = u(y,x) \Rightarrow [u(x,y) > u(y,y) \text{ or } \forall \varepsilon \in (0, \varepsilon^*) \text{ it holds that } y \notin \operatorname{argmax}_{z \in A} u(z, \varepsilon \cdot y + (1 - \varepsilon) \cdot x)]$.

Definition 3’ is equivalent to Definition 3. To see this, suppose first that strategy $x \in A$ is REE according to Definition 3. Swinkels (1992) shows that (x,x) then must be a (proper) Nash equilibrium, so $u(x,x) \geq u(y,x) \forall y \in A$, implying that x meets condition (i) of Definition 3’. That x also meets condition (ii) of Definition 3’ is immediate. Conversely, suppose $x \in A$ is REE according to Definition

3'. Given Swinkels (1992) finding that an ESS must be REE, it is easy to see that x then must be REE according to Definition 3.

Definition 3' makes clear how the REE and ESS criteria are related. Both criteria imply that for $x \in A$ to pass the respective tests, it must be a best response against itself. The ESS criterion moreover requires that any alternative best reply $y \in A$ to x must do strictly worse than x against y . This is sufficient for x to be REE, but x passes the test also if y is not an equilibrium entrant in Swinkels (1992) sense. Condition (ii) of Definition 3' effectively excludes as potential invaders all those alternative best responses to x that lack a quality of post-entry optimality.

In Swinkels (1992), p. 307, words, the REE criterion “strikes a middle ground between the total absence of rationality assumed by evolutionary game theory and the strenuous rationality requirements of traditional game theory”. However, it is not obvious that the only natural quality to impose on potential invaders is to assume they have the foresight to choose post-entry optimal strategies. A reasonable alternative might be to require entrants that are alternative best responses to x to choose strategies with some kind of decision-theoretic appeal. More generally, I suggest replacing conditions (ii) of Definition 1 to read ‘ $u(x,x) = u(y,x) \Rightarrow [u(x,y) > u(y,y) \text{ or } y \notin R]$ ’, where $R \subseteq A$ contains the set of strategies that are ‘rational’ according to some criterion.

Gans (1995) studies ‘evolutionary selection of beliefs’ and performs an exercise which can be interpreted within this framework. Translated to the present setting, he considers the cases where R is equal to respectively the set of strategies that are not strictly dominated and the set of rationalizable strategies (=those that survive iterated elimination of strictly dominated strategies). However, in these particular cases no leeway is offered in addition to Definition 1, so Gans (1995) concepts are equivalent to the ESS concept.¹

Now two new criteria will be considered. These relate to the following definition:

Definition 4. Let $A^0 = A$. For $t = 1, 2, \dots$, define A^t recursively by $A^t = \{x \in A^{t-1} \mid \text{there exists no } y \in A^{t-1} \text{ such that } u(y,z) \geq u(x,z) \forall z \in A^{t-1} \text{ and } u(y,z) > u(x,z) \text{ for some } z \in A^{t-1}\}$. A^1 is the set of admissible strategies. $\bigcap_{t=0}^{\infty} A^t$ is the set of iteratively admissible strategies.

According to Definition 4, a strategy is admissible iff 1 is weakly undominated. A strategy is iteratively admissible iff it survives iterated elimination of dominated strategies where at each round every dominated strategy is eliminated. It is well-known that (for the class of games under consideration) there exists $T \geq 0$ such that $\forall k \geq 0$ it holds that $\emptyset \neq A^{T+k} = \bigcap_{t=0}^{\infty} A^t$.

Definition 5. $x \in A$ is robust against admissible entrants (RAE) if $\forall y \in A, y \neq x$ it holds that (i) $u(x,x) \geq u(y,x)$, and (ii) $u(x,x) = u(y,x) \Rightarrow [u(x,y) > u(y,y) \text{ or } x \text{ is not admissible}]$.

Definition 6. $x \in A$ is robust against iteratively admissible entrants (RIE) if $\forall y \neq x$ it holds that (i) $u(x,x) \geq u(y,x)$, and (ii) $u(x,x) = u(y,x) \Rightarrow [u(x,y) > u(y,y) \text{ or } x \text{ is not iteratively admissible}]$.

In promoting his REE criterion, Swinkels (1992), Fig. 2, provides an example of a game in which no ESS exist and nevertheless a strategy (which is REE) seems ‘stable’. The reader may verify that

¹To see why this happens, suppose $x \in A$ satisfies either of Gans' criteria but yet is not ESS. Then there must exist a non-rationalizable strategy $y \in A$ such that $u(y,x) = u(x,x)$. However, since (x,x) must be a Nash equilibrium and any best reply to x then must be rationalizable it must hold that $u(y,x) < u(x,x)$. This is a contradiction.

strategy is also RAE and RIE. However, these two criteria are in general neither weaker nor stronger than the REE criterion. The game G_2 shows that neither of the two new criteria implies robustness against equilibrium entrants:

	<i>a</i>	<i>b</i>	<i>c</i>	
<i>a</i>	1,1	0,1	2,1	
<i>b</i>	1,0	1,1	0,1	
<i>c</i>	1,2	1,0	1,1	G_2

In G_2 , the strategy *a* is RAE and RIE but not REE (the dominated strategy *b* can invade).

On the other hand, robustness against equilibrium entrants does not imply any of the new criteria. This can be illustrated using the game G_3 :

	<i>r</i>	<i>s</i>	<i>p</i>	
<i>r</i>	1,1	2,0	0,2	
<i>s</i>	0,2	1,1	2,0	
<i>p</i>	2,0	0,2	1,1	G_3

In G_3 , the strategy $(r/3 + s/3 + p/3)$ is REE (cf. the discussion of Example 2.8 in Weibull (1995)), but this strategy is neither RAE nor RIE.

Neither of the two new criteria imply neutral stability. This can be shown using G_1 , in which game the strategy *a* is RAE and RIE but not NSS. Conversely, neutral stability does not imply any of the new criteria. This can be shown using a game such that $u(a,b) = u(c,d) \forall a,b,c,d \in A$ (with A not singleton), in which any strategy is NSS but no strategy is RAE or RIE.

In the games considered so far the sets of strategies which are RAE and RIE have coincided. It is obvious from the Definitions 5 and 6 that if $x \in A$ is RAE it must be RIE. The game G_4 illustrates that the converse is not true:

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	
<i>a</i>	1,1	0,1	0,1	2,1	
<i>b</i>	1,0	1,1	0,1	0,0	
<i>c</i>	1,0	1,0	1,1	0,1	
<i>d</i>	1,2	0,0	1,0	1,1	G_4

In G_4 , *a* (which is the unique iteratively admissible strategy) is RIE. However, *a* is not RAE (the admissible strategy *c* can invade).

It is well-known that strategies which are ES, NSS, or REE fail to exist in some games. The same goes for the RAE and RIE criteria. The game G_5 (from Ex. 2.7 in Weibull (1995)) even illustrates that the seemingly weak requirement that a strategy should meet *any* of the hitherto discussed criteria is incompatible with existence:

	<i>a</i>	<i>b</i>	<i>c</i>	
<i>a</i>	1,1	0,1	0,1	
<i>b</i>	0,1	1,1	1,0	
<i>c</i>	1,0	0,1	1,1	G_5

In G_5 , no strategy satisfies any of the criteria discussed so far (the successful invaders under the NSS criterion discussed in Ex. 2.7 in Weibull (1995) work for the other criteria too).

Finally, the game G_6 shows that a strategy may simultaneously satisfy all the criteria of NSS, REE, RIE, and RAE, but are not ESS:

	<i>a</i>	<i>b</i>	<i>c</i>	
<i>a</i>	1,1	1,1	1,1	
<i>b</i>	1,1	1,1	0,2	
<i>c</i>	1,1	2,0	0,0	G_6

In G_6 , the strategy a is not ESS (b can invade), but satisfies all the other properties.

4. Summary

Table 1 summarizes how the criteria defined in Section 2 and Section 3 are related. A ‘YES’ in some cell indicates that the criterion mentioned in the corresponding row implies the criterion mentioned in the corresponding column, etc. The bracketed information in the cell indicates which game or definition in this paper, or which other (abbreviated) reference, is relevant for understanding the result. For example, ‘NO [MS]’ in the NSS/ESS-cell means that neutral stability does not imply evolutionary stability, which is a result due to Maynard-Smith (1982).

Acknowledgements

I am indebted to Jörgen Weibull who has introduced me to evolutionary game theory and gotten me interested in the topic studied here. I thank a referee for helpful comments, and Handelsbankens forskningsstiftelser for financial support.

Table 1
Relationships between different stability criteria

	ESS	NSS	REE	RAE	RIE
ESS	YES [Def 1]	YES [MS]	YES [S]	YES [Def 5]	YES [Def 6]
NSS	NO [MS]	YES [Def 2]	NO [W]	NO [e.p.g.]	NO [e.p.g.]
REE	NO [S]	NO [G_1]	YES [Def 3]	NO [G_3]	NO [G_3]
RAE	NO [G_6]	NO [G_1]	NO [G_2]	YES [Def 5]	YES [Def 5]
RIE	NO [G_6]	NO [G_1]	NO [G_2]	NO [G_4]	YES [Def 6]

‘YES’ in a cell indicates that the row criterion implies the column criterion, etc. The bracketed information indicates which game, definition, or other (abbreviated) reference is relevant for the result.

‘e.p.g.’ means ‘equal payoffs game’ (any G with A not singleton and $u(a,b)=u(c,d) \forall a,b,c,d \in A$).

References

- Gans, J.S., 1995. Evolutionary selection of beliefs. *Economics Letters* 49, 13–17.
- Maynard-Smith, J., 1982. *Evolution and the Theory of Games*. Cambridge University Press.
- Maynard-Smith, J., Price, G.R., 1973. The logic of animal conflict. *Nature* 246, 15–18.
- Swinkels, J.M., 1992. Evolutionary stability with equilibrium entrants. *Journal of Economic Theory* 57, 306–332.
- Weibull, J.W., 1995. *Evolutionary Game Theory*. MIT Press.