Some relationships between evolutionary stability criteria in games

Martin Dufwenberg[1]

CentER for Economic Research, Tilburg University, The Netherlands
Department of Economics, Uppsala University, Sweden

(December, 1996)

Abstract: The relationships between five stability criteria for evolutionary games are studied.

Keywords: Evolutionary stability

JEL classification: C72

[1] Mailing address: Department of Economics, Uppsala University, Box 513, S-75120 Uppsala, Sweden; tel.: +46-18-182309; fax: +46-18-181478, e-mail: martin.dufwenberg@nek.uu.se. I'm indebted to Jörgen Weibull who has introduced me to evolutionary game theory and gotten me interested in the topic studied here, and I thank Handelsbankens forskningsstiftelser for financial support.

1. Introduction

I study the relationships between several "stability" criteria relevant to evolutionary game theory. The first results, presented in Section 2, concern some established concepts: I show that *neutral stability* (Maynard Smith 1982) does not imply *robustness against equilibrium entrants* (Swinkels 1992). It is already known that no implication runs the other way (Weibull 1995). I show also that the two criteria taken together do not imply *evolutionary stability* in the sense of Maynard Smith & Price (1973).

The second set of results, presented in Section 3, relate to two new stability criteria which I introduce: *robustness against admissible entrants* and *robustness against iteratively admissible entrants*. The former concept is strictly stronger than the latter. I argue that both have natural motivations in the spirit of Swinkels' (1992) criterion of robustness against equilibrium entrants. However, each of the new criteria neither implies nor is implied by respectively neutral stability and robustness against equilibrium entrants. I show also that there exist games in which no strategy satisfies any of the hitherto mentioned criteria. Finally, I show that the last result mentioned in the previous paragraph can be strengthened: Neutral stability taken together with robustness against both admissible and equilibrium entrants do not imply evolutionary stability.

The analysis below deals with mixed extensions 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 $u: A \times A \rightarrow \Re$ 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 $e \cdot y + (1-e) \cdot x$ is used to denote the strategy which puts probability weight e on y, etc.

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 forces:

DEFINITION 1 (Maynard Smith & Price 1973): $x \in A$ is an *evolutionarily stable strategy* (ESS) if $\forall y \in A$ 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 $\forall y \in A$ 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 $\exists e^* \in (0,1)$ such that $\forall y \in A$, $y \neq x$ and $\forall e \in (0,e^*)$ it holds that $y \notin argmax_{z \in A}\, u(z, e \cdot y + (1-e) \cdot x)$.

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 (1992, 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 following game 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

$\boldsymbol{G_1}$

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 $\forall e>0$ it holds that $y \notin argmax_{z \in A}\, u(z, e \cdot y + (1-e) \cdot a)$, 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 (based 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 $\exists e^* \in (0,1)$ such that $\forall y \in A$ it holds that (i) $u(x,x) \geq u(y,x)$, and (ii) $u(x,x)=u(y,x) \Rightarrow [u(x,y)>u(y,y)$ ***or*** $\forall e \in (0,e^*)$ it holds that $y \notin argmax_{z \in A}\ u(z, e \cdot y+(1-e) \cdot x)]$.

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 relate to each other. Both demand 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 exclude 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 of rationality assumed by evolutionary game theory and the strenuous rationality requirements of traditional game theory". However, it is not obvious that the only or the most 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 that satisfy some "rationality" properties that have some kind of decision-theoretic appeal. More generally, I suggest replacing condition (ii) of Definition 1 with a condition such that "$u(x,x)=u(y,x) \Rightarrow$

[$u(x,y)>u(y,y)$ ***or*** $y\notin R$]", where $R\subseteq A$ contains the set of strategies that are "rational" according to some criterion.

Gans (1994) studies "evolutionary selection of beliefs" and performs an exercise which fits into 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' (1994) concepts are equivalent to the ESS concept.[2]

Here, two new definitions will be considered. The following terminology is used: A strategy is *admissible* iff it is not (weakly) dominated, the strategy is *iteratively admissible* iff it survives iterated elimination of dominated strategies where at each round every dominated strategy is eliminated.

DEFINITION 4: $x\in A$ is *robust against admissible entrants* (RAE) if $\forall y\in A$ it holds that (i) $u(x,x)\geq u(y,x)$, and (ii) $u(x,x)=u(y,x) \Rightarrow$ [$u(x,y)>u(y,y)$ ***or*** x is not admissible].

DEFINITION 5: $x\in A$ is *robust against iteratively admissible entrants* (RIE) if $\forall y\in A$ it holds that (i) $u(x,x)\geq u(y,x)$, and (ii) $u(x,x)=u(y,x) \Rightarrow$ [$u(x,y)>u(y,y)$ ***or*** x 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 seems "stable". The reader may verify that that strategy is also RAE and RIE. However, these criteria are neither weaker nor stronger than the REE criterion. The following example based on the game G_2 shows that neither of the two new criteria implies robustness against equilibrium entrants:

---

[2] To 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 $z\in A$ such that $u(z,x)=u(x,x)$. However, since (x,x) must be a Nash equilibrium and any best reply to x then must be rationalizable it must hold that $u(z,x)<u(x,x)$. This is a contradiction.

	a	***b***	***c***	
a	1,1	0,1	2,1	
b	1,0	1,1	0,1	
c	1,2	1,0	1,1	$\boldsymbol{G_2}$

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 following game G_3:

	r	***s***	***p***	
r	1,1	2,0	0,2	
s	0,2	1,1	2,0	
p	2,0	0,2	1,1	$\boldsymbol{G_3}$

In G_3, the strategy ($r/3+s/3+p/3$) is REE (cf. the discussion of Example 2.8 in Weibull (1992)), 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 4 and 5 that if $x\in A$ is RAE it must be RIE. The following example illustrates that the converse is not true (in contrast to Gans' (1994) cases where iterated elimination of potential entrants makes no difference):

	a	***b***	***c***	***d***	
a	1,1	0,1	0,1	2,1	
b	1,0	1,1	0,1	0,0	
c	1,0	1,0	1,1	0,1	
d	1,2	0,0	1,0	1,1	$\boldsymbol{G_4}$

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 ESS, NSS, or REE fail to exist in some games. The same goes for the RAE and RIE criteria, as is evident from e.g. G_3. The following game (which appears in Example 2.7 in Weibull (1995)) illustrates that even the seemingly very weak stability requirement that a strategy should meet *any* of the hitherto discussed criteria would fail existence in some games:

	a	***b***	***c***	
a	1,1	1,0	0,1	
b	0,1	1,1	1,0	
c	1,0	0,1	1,1	$\boldsymbol{G_5}$

In G_5, no strategy satisfies any of the criteria discussed so far (the successful invaders under the NSS criterion discussed in Example 2.7 in Weibull (1992) work for the other criteria too).

Finally, the following game G_5 shows that a strategy may simultaneously satisfy all the criteria of NSS, REE, RIE, *and* RAE, but yet not be ESS:

	a	***b***	***c***	
a	1,1	1,1	1,1	
b	1,1	1,1	0,2	
c	1,1	2,0	0,0	$\boldsymbol{G_6}$

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

References

Gans, J., 1994, Evolutionary Selection of Beliefs, Discussion Paper 94/23 (The University of New South Wales).

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., 1992, Evolutionary Stability with Equilibrium Entrants, Journal of Economic Theory 57, 306-32.

Weibull, J., 1995, Evolutionary Game Theory (MIT Press).