

Lehrer, Ehud; Sorin, Sylvain

**Working Paper**  
**e-Consistent Equilibrium**

Discussion Paper, No. 1109

**Provided in Cooperation with:**

Kellogg School of Management - Center for Mathematical Studies in Economics and Management Science, Northwestern University

*Suggested Citation:* Lehrer, Ehud; Sorin, Sylvain (1994) : e-Consistent Equilibrium, Discussion Paper, No. 1109, Northwestern University, Kellogg School of Management, Center for Mathematical Studies in Economics and Management Science, Evanston, IL

This Version is available at:

<https://hdl.handle.net/10419/221465>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

Discussion Paper No. 1109

## $\epsilon$ -CONSISTENT EQUILIBRIUM

by

Ehud Lehrer\*

and

Sylvain Sorin\*\*

June 1994

Abstract: We deal with the concept of  $\epsilon$ -consistent equilibrium which corresponds to strategies inducing an  $\epsilon$ -equilibrium in any subgame reached along the play path. Examples and existence conditions are given.

---

\* Department of Managerial Economics and Decision Sciences, J.L. Kellogg Graduate School of Management and Department of Mathematics, Northwestern University, 2001 Sheridan Road, Evanston, Illinois 60208 and Raymond and Beverly Sackler Faculty of Exact Sciences, School of Mathematical Sciences, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel. e-mail: elehrer@casbah.acns.nwu.edu.

\*\* Département de Mathématiques et d'Informatique, URA CNRS 762, Ecole Normale Supérieure, 45 rue d'Ulm, 75230 Paris Cedex 05, France, and SEGMI, Université Paris X, 200 Ave. de la République, 92001 Nanterre, France; e-mail: sorin@dmi.ens.fr.

We thank Ehud Kalai for helpful discussions.

## 1. Introduction

In this paper we elaborate on the notion of  $\varepsilon$ -Nash equilibrium, where each player is optimizing up to an  $\varepsilon$ . Instead of playing a best response to the other player's strategies a player is adopting a strategy that can be improved upon. However, any improvement would guarantee an extra gain of at most  $\varepsilon$ .

There are three main justifications to  $\varepsilon$ -equilibrium. In Radner (1986) players have bounded computational capacity and therefore they cannot be fully rational. For computational reasons players can find only  $\varepsilon$ -optimizing strategies, rather than perfect best responses. In such a set up the best one can hope for is  $\varepsilon$ -equilibrium. Radner shows that  $\varepsilon$ -equilibrium allows for cooperation in a finitely repeated prisoners' dilemma.

The other two justifications involve infinitely repeated games. One way to think of infinite games is as an approximation of unspecified large finite games. Thus, an equilibrium of the infinite game is a profile which induces  $\varepsilon$ -equilibrium in any sufficiently long truncation. Therefore, the longer the game lasts the more precise the equilibrium gets. This is the uniform property introduced in Sorin (1990).

The game theoretical literature refers to processes that converge to equilibrium. Far enough in each one of these learning processes only  $\varepsilon$ -equilibrium is achieved and not full equilibrium: see, for instance, Kalai and Lehrer (1993).

Radner (1986) also mentions an elaborate definition of  $\varepsilon$ -equilibrium where players, at each stage of the game, are  $\varepsilon$ -rational. That is, players are consistent; they take into consideration the future payoffs they face and use  $\varepsilon$ -optimizing strategies at every period. This definition differs from the traditional definition in that in the latter there may exist a small event where players are not rational at all. Since such an event occurs with a small probability the overall effect on optimality is minor (i.e.,  $\varepsilon$ ). Thus, the traditional definition requires an  $\varepsilon$ -optimality only at the beginning of the game and

not during the game. The consistency property requires that the player will remain  $\varepsilon$ -rational all the way through.

To be  $\varepsilon$ -rational means that, whatever the history reached, as long as it is possible (i.e., having positive probability), each player is playing a best response up to an error of at most  $\varepsilon$ . Clearly, the magnitude of error depends on the payoff function defining the continuation game. In a discounted game, for instance, if the payoff function is not normalized and it is just the remaining payoff, then the  $\varepsilon$ -consistency requirement is vacuous in the long run. This is so because without normalization all payoffs are asymptotically less than  $\varepsilon$  and, therefore, any strategy is  $\varepsilon$ -consistent. In order for the  $\varepsilon$ -consistency requirement to bear some content, the payoff function should be defined in discounted games, as if at any time the game starts from the beginning.

Here, we introduce an  $\varepsilon$ -consistency requirement which may be reasonably applied to discounted games as well as to others. There is one instance where the  $\varepsilon$ -consistency property might create problems of nonexistence. This is the case of stochastic games. In the last section of this paper we consider stochastic games with absorbing states and show that  $\varepsilon$ -consistent equilibrium does exist.

## 2. $\varepsilon$ -Equilibrium

We first recall the traditional definition of  $\varepsilon$ -equilibrium in a strategic form game. Let player  $i$ 's set of actions be  $\Sigma^i$ . Set  $\Sigma = \times_i \Sigma^i$ . Player  $i$ 's payoff function is denoted by  $\gamma^i$ ;  $\gamma^i: \Sigma \rightarrow \mathbb{R}$ .

**Definition 1:** A profile  $\sigma \in \Sigma$  is an  $\varepsilon$ -equilibrium if for every player  $i$  and for all  $\tau^i \in \Sigma^i$

$$\gamma^i(\sigma^{-i}, \tau^i) \leq \gamma^i(\sigma) + \varepsilon.$$

Example 1: Consider now the following symmetric game, where  $\Sigma^i = [0,1]$ ,  $i = 1,2$ .

The payoff functions are defined as follows:

$$\gamma^1(s,t) = \begin{cases} 2t - s & \text{if } s > t \\ -1 & \text{if } s = t \\ s & \text{if } s < t \end{cases}, \quad \gamma^1(s,t) = \gamma^2(t,s).$$

In this example (1,1) is not 0-equilibrium (i.e., it is not a Nash equilibrium). However, (1,1) can be approximated by  $\varepsilon$ -equilibria. In such games the following definition makes sense.

Definition 2:  $x$  is an equilibrium payoff if for every  $\varepsilon > 0$  there exists an  $\varepsilon$ -equilibrium which induces a payoff within  $\varepsilon$  of  $x$ .

According to Definition 1 (1,1) is an equilibrium payoff of the previous example. The payoff (1,1) cannot be sustained by an exact equilibrium but it can be approximated up to any required accuracy by  $\varepsilon$ -equilibria. Thus, we refer to it as (ideal) equilibrium payoff.

In stochastic games one finds the same phenomenon. Even in zero-sum stochastic games exact equilibria may fail to exist. However, equilibrium payoffs do exist (see the "Big Match" game in Blackwell and Ferguson (1968)).

Another use of  $\varepsilon$ -equilibrium in defining an exact equilibria notion is the following. Consider a repeated game where  $\gamma_n^i$  denotes player  $i$ 's payoff in the  $n$ -fold repeated game.

Definition 3: Fix an  $\varepsilon > 0$ . The profile  $\sigma$  is a uniform  $\varepsilon$ -equilibrium if:

- (i) the payoffs  $\gamma_n(\sigma)$  converge to, say,  $\gamma(\sigma)$ ; and
- (ii) the gain from deviation is uniformly bounded by  $\varepsilon$ . That is,  $\exists N$  such that
 
$$\forall n \geq N, \forall \tau \in \Sigma, \forall_i \gamma_n^i(\sigma^{-i}, \tau^i) \leq \gamma_n^i(\sigma) + \varepsilon.$$

Now we use it in order to define uniform equilibrium.

**Definition 4:**  $\sigma$  is a uniform equilibrium sustaining the payoff  $x$  if  $\sigma$  is a uniform  $\varepsilon$ -equilibrium for every  $\sigma$  and, moreover,  $\gamma(\sigma) = x$ .

In other words, the same profile  $\sigma$  is adapted to any  $\varepsilon > 0$ , which may correspond to different (larger)  $n$ .

### 3. Consistency

#### 3a. The Definition

Definition 1 of  $\varepsilon$ -equilibrium allows small probability events where players do not act in a rational manner. In multistage games it means that there are decision nodes, reached with positive probability, where players do not optimize. Such a problem does not arise when dealing with exact equilibrium.

Consider an  $n$ -player multi-move strategic game  $G$  with payoff function  $\gamma = (\gamma^1, \dots, \gamma^n)$ . Denote by  $H$  the set of histories. Assume that each  $h$  corresponds to a subgame  $G(h)$  with payoff  $\gamma(h) = (\gamma^1(h), \dots, \gamma^n(h))$ .

**Definition 5:** A profile  $\sigma$  is an  $\varepsilon$ -consistent equilibrium if there exists a set  $P$  of paths having probability 1 according to  $\sigma$  satisfying:

- (C) For every  $h$  compatible with  $P$  (i.e., the set of those paths in  $P$  having  $h$  as their prefix is not empty), the induced strategy of  $\sigma$  in  $G(h)$ ,  $\sigma(h)$ , is an  $\varepsilon$ -equilibrium in  $G(h)$ .

In order to exemplify the role of  $\gamma(h)$ , let  $h$  be compatible with  $P$  and let  $h'$  be a history of  $G(h)$ . Thus,  $hh'$ , the concatenation of  $h$  and  $h'$ , is a history in  $G$ . In a discounted game, if  $\lambda^i$  is player  $i$ 's discount factor, then the natural way to define the payoff function in  $G(h)$  is  $\gamma^i(h)(h') = (\lambda^i)^{-|h|} \gamma^i(hh')$ , where  $|h|$  is the length of history  $h$ . The traditional way to define  $\gamma^i(h)(h')$  as  $\gamma^i(hh')$  is not satisfactory because it would make any  $\varepsilon$ -equilibrium  $\varepsilon$ -consistent (provided that  $\lambda^i$  is small enough).

A typical example of  $\varepsilon$ -consistent equilibrium in a repeated game is the stationary strategy which consists of playing repeatedly one-shot  $\varepsilon$ -equilibrium.

### 3b. Examples

We will now examine different classes of repeated games and their equilibria in regards to properties of the induced strategies.

In undiscounted infinitely repeated games with complete information the folk theorem can be stated as follows. For any feasible and individually rational payoff,  $x$ , there exists an equilibrium  $\sigma$  sustaining  $x$ . Moreover,  $\sigma$  is pure on the path it induces and  $\sigma(h)$  is an equilibrium sustaining  $x$  for any history having positive probability. In other words,  $\sigma$  induces in any subgame compatible with itself an equilibrium sustaining  $x$ . In fact, one could even obtain similar properties for perfect equilibrium.

In undiscounted infinitely repeated games with incomplete information on one side, for any equilibrium payoff  $x$  there exists an equilibrium sustaining it (because of the properties of Blackwell's (1956) approachability strategy and Hart's (1985) characterization of the set of equilibrium payoffs). Furthermore, after any history  $h$ ,

having positive probability,  $\sigma(h)$  is an equilibrium sustaining  $x(h)$ , which, due to the jointly controlled lottery conducted during the game, differs from  $x$  ( $x$  is the weighted averages of  $x(h)$  across histories of the same length).

In contrast to the previous two models of repeated games, in undiscounted stochastic games even in the zero-sum case an equilibrium sustaining the value generally fails to exist. However, uniform  $\varepsilon$ -equilibria that approximate the value does exist for every  $\varepsilon$ . In the non zero-sum case, existence of equilibrium payoff is known for 2 players game with absorbing states (Vrieze and Thuijsman (1989)). However, the strategies they constructed to prove this fact involve randomization and punishment on a set of path with positive probability. Thus, there is a positive probability set of histories along which the strategies defined are not optimal. In other words, the strategies Vrieze and Thuijsman defined are not  $\varepsilon$ -consistent.

The next section is devoted to the proof of existence of  $\varepsilon$ -consistent equilibria in this framework.

#### 4. Existence

The basic construction of Vrieze and Thuijsman's proof (see also Mertens, Sorin and Zamir (1994), pp. 406-408) is to have the players play stationary strategies and punish (forever) if after some stage the frequency of moves is too far from the reference strategy.

More precisely, there are two cases:

- (i) Players I and II play  $x$  and  $y$  i.i.d. where  $(x,y)$  is a nonabsorbing pair and any absorbing deviation is self punishing. If, at some stage  $n \geq N$ ,  $|\bar{x}_n - x| \geq \varepsilon$ , player II reduces player I's payoff to its max min, say,  $w^I$  (and similarly for  $\bar{y}_n$  and  $w^{II}$ ).
- (ii) Player I plays  $x$  i.i.d. and player II plays an i.i.d. mixture  $(1 - \varepsilon, \varepsilon)$  of  $y$  and



$z$  with  $(x,y)$  nonabsorbing and  $(x,z)$  absorbing. Any absorbing deviation of player II (versus  $x$ ) is self punishing, as well as any absorbing deviation of player I versus  $y$ . As above, player II punishes player I if after some large stage  $N_1$  the frequency is far from  $x$ , while player I punishes player II from some stage  $N_2$  on (since basically the game should be over before).

Assume that all payoffs are bounded in absolute value by one. Let us denote by  $\sigma$  a strategy of player I satisfying, given  $\delta > 0$ , the following conditions:

- (1)  $\exists N_0$ , such that for all  $n \geq N_0$ , the expected average player II's payoff up to stage  $n$  is less than  $w^{II} + \delta$  and so is the absorbing payoff at stage  $n$ , if the total probability of absorption at stage  $n$  is more than  $\delta$ .
- (2) Given any history  $h$ , there exists a move  $j(h)$  of player II such that the conditional probability of absorption (i.e., the probability of absorption at this stage conditionally to nonabsorption until that stage) against  $\sigma(h)$  is less than  $\delta$ .

We will now choose  $N_1$  such that the probability that by playing  $y$  i.i.d. the event  $A = \{|\bar{y}_n - y| \geq \delta^2\}$  occurs for some  $n \geq N_1$  is less than  $\delta$ .

Let us first consider case (i).

After  $N_1$  stages, if  $A$  occurs for the first time at stage  $n_1$ , player I uses  $\sigma$  until some stage  $p_1$  which is either  $n_1 + N_2$  (where  $N_2$  is greater than  $N_0$  and  $N_1/\delta$ ) or the first stage where the conditional probability of absorption since stage  $n_1$  exceeds  $\delta$ , if it occurs before stage  $n_1 + N_2$ . Player II uses  $j(h)$  at each stage between  $n_1$  and  $p_1$ . We refer to the period before  $n_1$  as the first regular block and call it short if  $n_1 = N_1$  and long otherwise. Similarly, the period between  $n_1$  and  $p_1$  where  $\sigma$  is used (it is the first punishment block) is called absorbing if  $p_1 < n_1 + N_2$  and long otherwise. Then player I starts again playing  $x$  i.i.d. and computing frequency from stage  $p_1$  on and similarly for

player II.

If for the second time at some stage  $n_2 \geq p_1 + N_1$  the event A occurs--namely, the average of the moves of player II between stages  $p_1$  and  $n_2$  differs from  $y$  by more than  $\delta$ --player I uses again  $\sigma$ , but taking as initial history the history  $h_1$  corresponding to the first punishment block. Again this lasts for  $N_2$  stages unless the conditional absorption probability since stage  $n_2$  reaches  $\delta$  and then player I switches once again to  $x$ , and so on.

One defines thus inductively punishment blocks where  $\sigma$  is used, taking as past history at the beginning of a new block the concatenation of all past histories along all previous punishment blocks.

We claim that this strategy and a comparable one for player II, induce an  $\varepsilon$ -equilibrium, for  $\varepsilon = 6\delta$ .

Formally we introduce similarly the event B and the strategy  $\tau$  and use a lexicographic order, during the regular block to check the deviations, i.e., the occurrence of A or B.

Define  $K_1$  such that  $1 - (1 - \delta)^{K_1} \geq 1 - \delta$ , let then  $K_2 \geq K_1/\delta$  and  $K = K_1 + K_2$ .

Then there exists  $\bar{N}$  such that, if for some  $n \geq \bar{N}$ , the frequency  $\bar{y}_n$  is at a distance of at least  $\delta$  from  $y$ , then the event A occurred before stage  $n$  at least  $K$  times.

Now either there were at least  $K_1$  absorbing punishment blocks and the game is over with probability at least  $1 - \delta$ , or the number of long punishment blocks is at least  $K_2$ . Obviously the average nonabsorbing payoff corresponds to a convex combination of payoffs during punishing blocks and payoffs during regular blocks. Now either a regular block is small or the payoff excepted for one stage (the last one) is compatible with a frequency within  $\delta$  of  $y$ . Furthermore, the relative weight of the small blocks not followed by a long punishment is at most  $\delta$ . A similar statement applies also to the

weight of the absorbing punishment blocks with the difference that, on those, the average payoff is within  $2\delta$  of  $y$ .

In both cases (whether there were at least  $K_1$  absorbing punishments or less), the expected absorbing payoff--obtained during all punishment phases--is less than the  $\max \min + \delta$  if the probability of absorption at stage  $n$  is more than  $\delta$ .

The strategies just described are  $\varepsilon$ -consistent. In fact on any nonabsorbing history the expected payoff, say, after a deviation of player II, is within  $\varepsilon$  of the original payoff: the expected payoff after the current punishment is exactly the initial one and the probability of absorption is at most  $2\delta$ .

Finally, concerning case (ii), the behavior of player II will be the same as above, while player I will consider regular blocks of fixed size  $M$  (such that under  $x$  and  $(y,z)$  the probability of absorption exceeds  $1 - \delta$ ) and otherwise use  $\sigma$  for  $N_1$  stages as above.

References

- Blackwell, D. (1956), "An Analog of the Minmax Theorem for Vector Payoffs," Pacific Journal of Mathematics, 6, 1-8.
- Blackwell, D. and T. S. Ferguson (1968), "The 'Big Match'," Annals of Mathematical Statistics, 39, 159-164.
- Hart, S. (1985), "Nonzero-sum Two-person Repeated Games With Incomplete Information," Mathematics of Operations Research, 10, 117-153.
- Kalai, E. and E. Lehrer (1993), "Rational Learning Leads to Nash Equilibrium," Econometrica, 61, 1019-1045.
- Radner, R. (1986), "Can Bounded Rationality Resolve the Prisoners' Dilemma?" in Contributions to Mathematical Economics in Honor of G. Debreu, W. Hildenbrand and A. Mas-Colell (eds.), North Holland, 387-399.
- Mertens, J.-F., S. Sorin and S. Zamir (1994), Repeated Games, CORE D.P. No. 9420-21-22.
- Sorin, S. (1990), "Supergames," in Game Theory and Applications, T. Ichiichi, A. Neyman and Y. Taumann (eds.), pp. 46-63, Academic Press.
- Vrieze, O. J. and F. Thuijsman (1989), "On Equilibria in Repeated Games with Absorbing States," International Journal of Game Theory, 18, 293-310.