

Tilburg University

## History-dependent equilibrium points in dynamic games

van Damme, E.E.C.

*Published in:*  
Game theory and mathematical economics

*Publication date:*  
1981

*Document Version*  
Publisher's PDF, also known as Version of record

[Link to publication in Tilburg University Research Portal](#)

*Citation for published version (APA):*  
van Damme, E. E. C. (1981). History-dependent equilibrium points in dynamic games. In O. Moeschlin, & D. Pallaschke (Eds.), *Game theory and mathematical economics: Proceedings of the Seminar on game theory and mathematical economics, Bonn/Hagen, 7-10 October, 1980* (pp. 27-38). North-Holland Publishing Company.

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# History-dependent equilibrium points in dynamic games

Damme, van, E.E.C.

Published: 01/01/1980

## *Document Version*

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

### **Please check the document version of this publication:**

- A submitted manuscript is the author's version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## *Citation for published version (APA):*

Damme, van, E. E. C. (1980). History-dependent equilibrium points in dynamic games. (Memorandum COSOR; Vol. 8005). Eindhoven: Technische Hogeschool Eindhoven.

## **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

## **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

EINDHOVEN UNIVERSITY OF TECHNOLOGY

Department of Mathematics

PROBABILITY THEORY, STATISTICS AND OPERATIONS RESEARCH GROUP

Memorandum COSOR 80-05

History-dependent equilibrium points  
in dynamic games

by

E.E.C. van Damme

Eindhoven, May 1980

The Netherlands

# History-dependent equilibrium points in dynamic games

by

E.E.C. van Damme

## Abstract

A dynamic system, which is observed (and influenced) by two players simultaneously at discrete points in time, is considered. Both players receive rewards depending on the time path of the system and on the actions taken. The game is played noncooperatively. At an observation point a player can base his decision on the state of the system at that moment (memoryless strategy) or on the entire history of the system (history-dependent strategy). Some examples are given of games in which there exists an equilibrium point in history-dependent strategies, which gives both players a greater reward than any equilibrium point in memoryless strategies does. When the players play according to such a history-dependent equilibrium point, they make implicit agreements and hence cooperate implicitly. These implicit agreements attain stability by the use of threats. Although only some examples are given, it will become clear that the phenomenon is present in a wide class of dynamic games.

## 1. Introduction

In this paper we will be concerned with dynamic 2-person noncooperative non-zero-sum games. That the game is noncooperative means that binding agreements between the players are not possible.

For such games a Nash equilibrium point (= E.P.) constitutes a reasonable solution concept, although many criticisms against it have been raised. However, a great deal of this criticism was based on the fact that the concept yields counterintuitive solutions for some static games. But, since most of the real life conflict situations are dynamic in nature, we actually should investigate what solutions the concept yields for dynamic games.

Most of the present literature on noncooperative non-zero-sum dynamic games (e.g. Markov Games, Differential Games) deals with questions as:

- i) Does an E.P. exist?
- ii) If an E.P. exists, is it unique?
- iii) Are there E.P.'s which have a special structure (e.g. memoryless, or linear in the state)?

However, the question "how reasonable" an E.P. is, usually remains unanswered. By "how reasonable" we mean aspects as:

- iv) How high are the E.P.-rewards, compared with the rewards associated with other reasonable strategy pairs ?
- v) How likely is it that the E.P. will prevail in the actual playing of the game ?

The answers to the latter questions are not in favour of the E.P.'s investigated up to now (see e.g. example 5 where the results are very counterintuitive). However, up to now, usually only memoryless strategies have been considered. In this paper we will show that this restriction is not always justified. We will give a few examples of dynamic games, in which there exists an E.P. in history-dependent strategies, which is for both players better than any E.P. in memoryless strategies. From the examples it will become clear that the phenomenon is present in a wide class of dynamic games, not only 2-person, but also n-person games. Furthermore, in the examples the history-dependent E.P.'s are intuitively very appealing. So the phenomenon can be helpful to explain real life behaviour.

In general there is not just one E.P. in history-dependent strategies, which is for both players better than any E.P. in memoryless strategies. This makes the choice between different "good" E.P.'s a difficult one. In our opinion this choice problem is a problem of cooperative game theory, rather than of noncooperative game theory. We will look into this problem in a forthcoming paper.

## 2. The basic model

We will use  $P^i$  as an abbreviation of "player i" ( $i \in \{1, 2\}$ ). Let  $T$  be the set of relevant time points,  $T = \{1, \dots, T\}$  and let  $\Theta = T \setminus \{T\}$ .

Let  $X$  be a finite set, the state space. Let  $U$  and  $V$  be finite sets, the

action spaces of  $P^1$  and  $P^2$ , respectively. Let  $U(V)$  be the set of all probability distributions on  $U(V)$ . For  $p \in U$ , let  $C^1(p) := \{u \in U : p(u) > 0\}$ .  $C^2(q)$  is defined similarly, if  $q \in V$ . At each time point  $t \in \Theta$  both players make an observation on a dynamic system and depending upon the observed state they choose an action (possibly by using a random device) in order to influence the system. If at time  $t$  the system is in state  $x(t)$  and if actions  $u(t)$  and  $v(t)$  are taken then at time  $t+1$  the system is in

$$x(t+1) = f_t(x(t), u(t), v(t)).$$

Furthermore, in this situation  $P^1$  receives a reward  $r_t^1(x(t), u(t), v(t))$  at time  $t$ . In addition  $P^1$  receives a terminal reward  $r_T^1(x(T))$ . We assume that each player has perfect recall on the history of the system. This means that at time  $t$  ( $t \in \Theta$ ) a player can base his decision upon all the states he has observed up to and including time  $t$  and all the actions he and the other player have taken up to time  $t$ . The reader might argue that it would be more realistic to assume that at time  $t$  a player does not necessarily know exactly what actions the other player has taken up to time  $t$ , but only knows the rewards he received up to time  $t$ . However, in our examples a player can deduce from the state at time  $t$ , the action he has taken at time  $t$ , the reward he received at time  $t$  and the state at time  $t+1$ , what action the other player has taken at time  $t$ . So our assumption is not too restrictive. So the information available to a player at time  $t$  is an element of

$$H_t := (X \times U \times V)^{t-1} \times X.$$

A strategy for  $P^1$  is a sequence  $\pi = \{\pi_t\}_{t \in \Theta}$  such that  $\pi_t : H_t \rightarrow U$  ( $t \in \Theta$ ). A strategy  $\pi$  is called a *memoryless strategy* if it satisfies for all  $t \in \Theta$ :

$$\text{if } h, h' \in (X \times U \times V)^{t-1}, x \in X, \text{ then } \pi_t(h, x) = \pi_t(h', x).$$

So, actually a memoryless strategy for  $P^1$  is a sequence  $\pi = \{\pi_t\}_{t \in \Theta}$ , such that  $\pi_t : X \rightarrow U$  ( $t \in \Theta$ ).  $S^1(M^1)$  is the set of all (memoryless) strategies of  $P^1$ .  $S^2$  and  $M^2$  are defined similarly. We will use  $\sigma$  as a notation for a strategy of  $P^2$ . For  $\pi \in S^1$ ,  $\sigma \in S^2$  we define inductively:

$$H_1(\pi, \sigma) = X$$

$$H_{t+1}(\pi, \sigma) =$$

$$\{(h, x, u, v, f_t(x, u, v)); (h, x) \in H_t(\pi, \sigma), u \in C^1(\pi_t(h, x)), v \in C^2(\sigma_t(h, x))\}$$

So,  $H_t(\pi, \sigma)$  represents the set of all information a player could have at time  $t$ , when the strategies  $\pi$  and  $\sigma$  are played.

We will use  $R^i(x_1, \pi, \sigma)$  to denote the expected total reward for  $P^i$ , given that the initial state is  $x_1$  and the strategies  $\pi$  and  $\sigma$  are played. So

$$R^i(x_1, \pi, \sigma) := \mathbb{E}_{x_1, \pi, \sigma} \left\{ \sum_{t=1}^{T-1} r_t^i(x(t), u(t), v(t)) + r_T^i(x(T)) \right\}.$$

With these elements a game  $\Gamma(x) := \langle S^1, S^2, R^1(x, \dots), R^2(x, \dots) \rangle$  is defined for all

$x \in X$ . Let  $\Gamma := (\Gamma(x))_{x \in X}$ . A strategy pair  $(\hat{\pi}, \hat{\sigma})$  is a *Nash equilibrium point* (= E.P.) of  $\Gamma$  if, for all  $x \in X, \pi \in S^1, \sigma \in S^2$ :

$$R^1(x, \pi, \hat{\sigma}) \leq R^1(x, \hat{\pi}, \hat{\sigma})$$

$$R^2(x, \hat{\pi}, \sigma) \leq R^2(x, \hat{\pi}, \hat{\sigma}).$$

Let us denote by  $\Gamma^t(x_t)$  the subgame of  $\Gamma$ , which starts at time  $t$  in state  $x_t$ . Let  $(\hat{\pi}, \hat{\sigma})$  be a pair of memoryless strategies.  $(\hat{\pi}, \hat{\sigma})$  is a *recursive equilibrium point* (= R.E.P.) of  $\Gamma$  if for all  $t \in \Theta$  and all  $x_t \in X$  that can be reached at time  $t$  by some strategy pair  $(\pi, \sigma)$  we have:

$$(\hat{\pi}_t, \hat{\sigma}_t)_{t=T}^{T-1} \text{ is an E.P. of the game } \Gamma(x_t).$$

So a R.E.P. can be found by using a dynamic programming technique (see [3], also compare [6]).

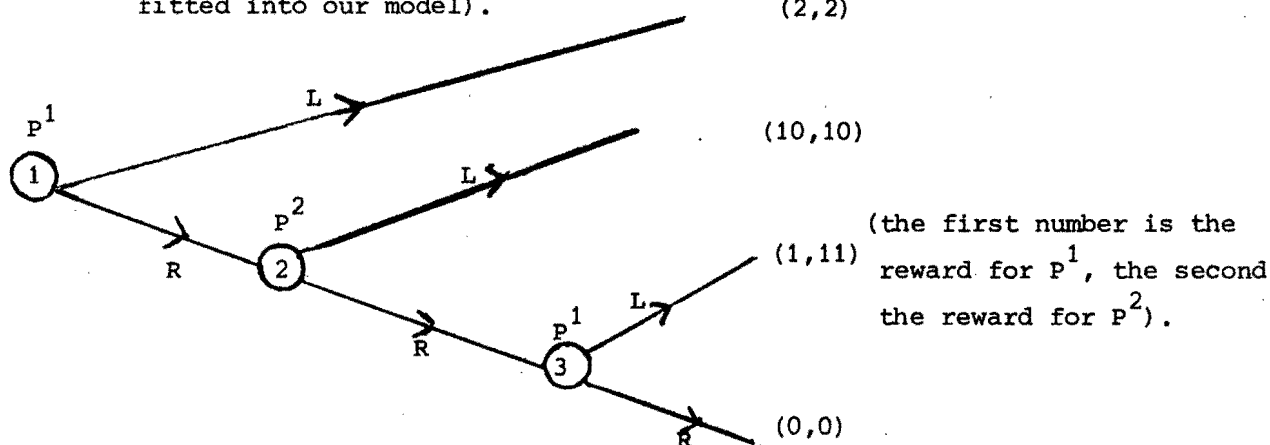
For purposes of clarity we give the following lemma:

**Lemma 1** Let  $\Gamma$  be a game defined as above. Let  $\Gamma|_M$  be the game in which  $P^1$  only uses strategies from  $M^1$  and  $P^2$  only uses strategies from  $M^2$ .

- i) If  $(\pi, \sigma)$  is an E.P. of  $\Gamma|_M$ , then  $(\pi, \sigma)$  is an E.P. of  $\Gamma$ .
- ii) If  $(\pi, \sigma)$  is a R.E.P. then  $(\pi, \sigma)$  is an E.P.
- iii) If  $(\pi, \sigma)$  is an E.P. and  $\pi \in M^1, \sigma \in M^2$ , then it is not necessarily the case that  $(\pi, \sigma)$  is a R.E.P.

Proof:

- i) Assume  $P^2$  uses strategy  $\sigma$  ( $\sigma \in M^2$ ) in  $\Gamma$ . Then the resulting problem for  $P^1$  is actually a standard deterministic dynamic programming problem, for which we know that there exists an optimal strategy, that is memoryless (see e.g. [5], lemma 1).
- ii) Trivial.
- iii) Consider the following game in extensive form:  
(the reader will have no difficulty in verifying that this game can be fitted into our model).



The R.E.P. is in this case:

$P^1$  plays L in state 1, L in state 3,  
 $P^2$  plays R in state 2.

For this game another E.P. in memoryless strategies is:

$P^1$  plays R in state 1, R in state 3,  
 $P^2$  plays L in state 2.

□

Although lemma 1 shows, that we have to distinguish between E.P.'s in memoryless strategies and R.E.P.'s, it is easy to show that in all the examples from now on these two concepts coincide. So, in the examples E.P.'s in memoryless strategies can be found by dynamic programming.

### 3. Repeated bimatrix games

Let  $(A, B)$  be a pair of  $m \times n$  matrices. The bimatrix game  $G = (A, B)$  is played as follows:

$P^1$  ( $P^2$ ) chooses (possibly by using a random mechanism) a row (a column). If row  $k$  and column  $l$  are chosen, then  $P^1$  receives a reward  $a_{kl}$  and  $P^2$  receives  $b_{kl}$ . In the following we are going to play such a game repeatedly. Each time



the same bimatrix game is played. The players try to maximize the sum of their payoffs. It is clear that such a game fits into our basic model.

Example 1 repeated prisoners' dilemma.

Let  $G$  be the following bimatrix game

$$G = \begin{array}{c|cc} & V_1 & V_2 \\ \hline U_1 & (10,10) & (0,11) \\ U_2 & (11,0) & (1,1) \end{array}$$

So, if  $P^1$  chooses  $U_1$  and  $P^2$  chooses  $V_2$ , then  $P^1$  gets a reward of 0 and  $P^2$  receives 11. If  $G$  is played once, there is just one E.P., namely  $(U_2, V_2)$ . The point  $(U_1, V_1)$  is better for both players, but since they are both motivated to shift away from it, they cannot reach it. Now let  $G(T)$  be the game in which  $G$  is played  $T$  times in succession. Intuitively one would expect that whenever  $T$  is large enough, it will be possible for the players to reach  $(U_1, V_1)$ . Namely, one would expect that in a dynamic context the players will make implicit agreements to reach points which are favourable for both, when explicit ones are forbidden. However, it turns out that this is not possible; all E.P.'s result in a repeatedly playing of  $(U_2, V_2)$ . A proof of this can already be found in Luce and Raiffa ([7]); however, this proof is not entirely correct (see example 3). Therefore we will present a proof in proposition 1.

Proposition 1 Let  $G$  be the game from example 1. Let  $G(T)$  be defined as above. Let  $(\pi, \sigma)$  be a strategy pair for  $G(T)$  and let  $H_t(\pi, \sigma)$  ( $t \in \{1, \dots, T\}$ ) be defined as in section 2. We have:

If  $(\pi, \sigma)$  is an E.P. in  $G(T)$ , then

$$\pi_t(h) = U_2 \text{ and } \sigma_t(h) = V_2,$$

for all

$$h \in H_t(\pi, \sigma) \text{ and } t \in \{1, \dots, T\}.$$

Proof The proof is given by backward induction.

The assertion of the theorem is true for  $t = T$ .

Let  $t < T$ ,  $\hat{h} \in H_t(\pi, \sigma)$ . Assume  $\pi_t(\hat{h}) = p \cdot U_1 + (1 - p) \cdot U_2$  ( $p > 0$ ).

By induction hypothesis we have:

$$\sigma_s(h) = V_2 \text{ for all } h \in H_s(\pi, \sigma) \text{ and all } s > t.$$

From this it follows:

$$r^1(A_2, \sigma_s(h')) \geq r^1(\pi_s(h), \sigma_s(h)),$$

for all

(\*)

$$h \in H_s(\pi, \sigma), h' \in H_s \text{ and } s > t.$$

So, if at some moment before  $t+1$   $P^1$  has deviated from  $\pi$ , then from  $t+1$  on, he is not worse off in this new situation than he was, when he had not deviated. Now, let the strategy  $\tilde{\pi}$  for  $P^1$  be defined by:

$$\tilde{\pi}_s(h) = \pi_s(h) \text{ for all } h \in H_s, s < t,$$

$$\tilde{\pi}_s(h) = A_2 \text{ for all } h \in H_s, s \geq t.$$

When at time  $t$  the information  $\hat{h}$  occurs, then  $\tilde{\pi}_t$  gives a greater reward than  $\pi_t$  does. Since  $\hat{h} \in H_t(\pi, \sigma)$  we have that the probability that  $\hat{h}$  occurs is positive.

Now  $\tilde{\pi}$  gives the same reward as  $\pi$  up to time  $t$  and (by(\*)) does not lose anything from time  $t+1$  on. So  $\tilde{\pi}$  is a better reply against  $\sigma$  than  $\pi$  is. This is a contradiction.

□

#### Remark

Reviewing the proof of proposition 1 we see that in the general case it can go wrong at two essential points:

i)  $T = \infty$  so that we cannot start the induction.

We will look into this case in example 2.

ii) The inequality in (\*) may not be true. In this case a deviation of  $P^1$  causes a deviation of  $P^2$  at some later time, such that the net result for  $P^1$  is negative. For this case, see example 3.

#### Example 2 infinitely often repeated prisoners' dilemma.

Let  $G$  be the game of example 1. When we consider an infinite repetition of  $G$ , we have that the total reward is not a well defined criterion to compare different strategies. Therefore, we will assume that the players use the average payoff per play as their performance criterion. So let

$$R_{\infty}^i(\pi, \sigma) := \liminf_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T r^i(\pi_t, \sigma_t)$$

Now let  $G(\infty)$  be the infinite repetition of  $G$ . In proposition 2 we show that in  $G(\infty)$  it is possible for the players to reach a better performance by using history-dependent strategies.

Proposition 2 Let  $G(\infty)$  be the game of example 2.

i) If  $(\pi, \sigma)$  is an E.P. such that  $\pi$  and  $\sigma$  are both memoryless strategies, then

$$R_{\infty}^1(\pi, \sigma) = 1 \text{ and } R_{\infty}^2(\pi, \sigma) = 1$$

ii) There exists an E.P.  $(\pi, \sigma)$  such that

$$R_{\infty}^1(\pi, \sigma) = 10 \text{ and } R_{\infty}^2(\pi, \sigma) = 10 .$$

Proof

i) Assume  $(\pi, \sigma)$  is an E.P., such that  $\pi \in M^1, \sigma \in M^2$ .

Assume  $\pi_t(G) = p_t \cdot U_1 + (1 - p_t) \cdot U_2$  and  $\sigma_t(G) = q_t \cdot V_1 + (1 - q_t) \cdot V_2$ .

It is not difficult to verify that  $P^1$  ( $P^2$ ) cannot play  $U_2$  ( $V_2$ ) too often. More precisely:

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T p_t = 0 \text{ and } \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T q_t = 0 .$$

From this, the statement follows immediately.

ii) If  $t \in \mathbb{N}$ , let  $\hat{h}_t \in H_t$  be defined by

$$\hat{h}_1 = G, \hat{h}_{t+1} = (\hat{h}_t, U_1, V_1, G)$$

Let the strategies  $\pi$  and  $\sigma$  be defined by

$$\pi_t(h_t) = \begin{cases} U_1 & , \text{ if } h_t = \hat{h}_t \\ U_2 & , \text{ if } h_t \neq \hat{h}_t \end{cases} \quad (t \in \mathbb{N})$$

$$\sigma_t(h_t) = \begin{cases} V_1 & , \text{ if } h_t = \hat{h}_t \\ V_2 & , \text{ if } h_t \neq \hat{h}_t \end{cases} \quad (t \in \mathbb{N})$$

So  $P^1$  ( $P^2$ ) plays  $U_1$  ( $V_1$ ) as long as  $P^2$  ( $P^1$ ) plays  $V_1$  ( $U_1$ ). If  $P^2$  shifts to  $V_2$  ( $U_2$ ) then  $P^1$  ( $P^2$ ) will play  $U_2$  ( $V_2$ ) for ever on. It can easily be verified that the pair  $(\pi, \sigma)$  is indeed an E.P. This E.P. yields both players a payoff of 10. So, it is for both better than any E.P. in memoryless strategies.  $\square$

Remark

Although the E.P. of proposition 2-ii) is a very nice E.P., it is by no means the only one that results in a better performance for both players than the E.P.'s of proposition 2-i). Namely, consider e.g. the case in which the players make the following "gentlemen's agreement": we start the game with playing  $N_1$  times  $(U_1, V_1)$ . After this we play  $N_2$  times  $(U_1, V_2)$ , then  $N_3$  times  $(U_2, V_1)$ , then  $N_4$  times  $(U_2, V_2)$ . Then again we play  $N_1$  times  $(U_1, V_1)$ , etc. Furthermore  $P^1$  ( $P^2$ ) decides that he will play  $U_2$  ( $V_2$ ) from the moment  $P^2$  ( $P^1$ ) has broken the agreement. Such a pair of strategies forms an E.P. as long as

$$\frac{10N_1 + 11N_3 + N_4}{\sum_{i=1}^4 N_i} \geq 1 \quad \text{and} \quad \frac{10N_1 + 11N_2 + N_4}{\sum_{i=1}^4 N_i} \geq 1$$

Now, let  $(x, y) \in \text{ch} \{ (0, 11), (11, 0), (10, 10), (1, 1) \}$ , (\*) such that  $x \geq 1$ ,  $y \geq 1$ . The reader will have no difficulty in constructing a pair of strategies  $(\pi, \sigma)$ , such that

- i)  $(\pi, \sigma)$  is an E.P. of  $G(\infty)$ .
- ii)  $R_\infty^1(\pi, \sigma) = x$  and  $R_\infty^2(\pi, \sigma) = y$ .

So, there is a great similarity between the static cooperative version of  $G$  and  $G(\infty)$ . For this result also see Aumann ([1]).

Example 3 Let  $G'$  be the following bimatrix game:

	$V_1$	$V_2$	$V_3$
$U_1$	(10, 10)	(0, 11)	(-5, -5)
$U_2$	(11, 0)	(1, 1)	(-4, -5)
$U_3$	(-5, -5)	(-5, -4)	(-5, -5)

$G'$  is almost the same game as  $G$  (remark that  $U_2$  ( $V_2$ ) strictly dominates  $U_1$  and  $U_3$  ( $V_1$  and  $V_3$ )). However, now each player has an extra action available which causes both players much harm. When this game is played repeatedly it is possible for the players to reach the point  $(U_1, V_1)$  by using history-dependent strategies. The availability of  $U_3$  ( $V_3$ ) gives  $P^1$  ( $P^2$ ) the possibility

---

((\*)  $\text{ch}(S)$  denotes the convex hull of  $S$ )

to force  $P^2$  ( $P^1$ ) to play  $V_1$  ( $U_1$ ). Namely, when  $P^2$  ( $P^1$ ) does not cooperate,  $P^1$  ( $P^2$ ) will punish him by playing  $U_3$  ( $V_3$ ). The reader can easily verify that the pair  $(\pi, \sigma)$  with

$$\begin{aligned}\pi_1 &= U_1 \\ \pi_2 &= \begin{cases} U_2 & \text{if } P^2 \text{ has played } V_1 \text{ the first time} \\ U_3 & \text{otherwise} \end{cases} \\ \sigma_1 &= V_1 \\ \sigma_2 &= \begin{cases} V_2 & \text{if } P^1 \text{ has played } U_1 \text{ the first time} \\ V_3 & \text{otherwise} \end{cases}\end{aligned}$$

is an E.P. in  $G'(2)$

This E.P. gives both players a payoff of 11. Furthermore, it is easy to see that the point  $(U_1, V_1)$  can only be reached by using history-dependent strategies. Namely, if  $(\pi, \sigma)$  is an E.P. of  $G'(2)$  and  $\pi \in M^1$ ,  $\sigma \in M^2$ , then:

$$\begin{aligned}\pi_1 &= U_2 & \sigma_1 &= V_2 \\ \pi_2 &= U_2 & \sigma_2 &= V_2\end{aligned}$$

This E.P. gives both players a payoff of 2. So, it is worthwhile to consider history-dependent strategies.

One might have the opinion that it is not very likely that the history-dependent equilibrium point of example 3 will actually be played. However, if we modify the game slightly (as in example 4) this becomes very likely.

#### Example 4

		$V_1$	$V_2$
$G'' =$	$U_1$	(10, 10)	(0, 11)
	$U_2$	(5, 0)	(1, 1)
	$U_3$	(5, -10)	(1, -9)

We consider  $G''(2)$  (the game in which  $G''$  is played twice in succession). In this game all E.P.'s in memoryless strategies result in a payoff of 2 for  $P^1$  and a maximum payoff of 2 for  $P^2$ . However, there exists an E.P. (in which  $P^1$  uses a history-dependent strategy) which gives both payers a payoff of 11. Namely let

$$\begin{aligned}\pi_1 &= U_1 \\ \pi_2 &= \begin{cases} U_2 & \text{if } P^2 \text{ played } V_1 \text{ the first time} \\ U_3 & \text{otherwise} \end{cases} \\ \sigma_1 &= V_1 \\ \sigma_2 &= V_2\end{aligned}$$

Then  $(\pi, \sigma)$  is the desired E.P.

In this section we have seen that in dynamic games we cannot restrict ourselves to memoryless strategies. However, our examples were a little artificial. In the next two sections we show that the same is the case for more realistic games. Section 4 gives a (simplified) example of an economic situation and section 5 considers a class of games that are actually used as a model of the decision process in the European Common Market ([8]).

#### 4. A market situation

Consider the following market situation:

2 sellers ( $P^1$  and  $P^2$ ) sell the same product. There are  $M$  buyers, who will each buy exactly one unit of the product. Both sellers charge the same price,  $p$  per unit of the product. This price cannot be changed. Each day the sellers decide whether they will advertise that day or not. Each day at most one advertisement can be made, this costs  $c$ . Depending on the actions taken, the number of customers for that day for each of the sellers is given in the following table:

$P^1 \backslash P^2$	A	N	
A	(3, 3)	(4, 1)	(A: advertise) (N: not advertise)
N	(1, 4)	(2, 2)	

We assume that, if some day  $P^1$  takes action A and  $P^2$  takes action N, while the number of people who have not bought the product yet (denoted by  $k$ ) is less than 5, then the expected number of customers for  $P^1$  ( $P^2$ ) is  $\frac{4}{5}k$  ( $\frac{1}{5}k$ ). A similar assumption is made in all other cases where the total number of customers as indicated in the table exceeds the number of people who have not bought the product yet.

This situation can be fitted into our basic model:

Let  $X = \{0, 1, \dots, M\}$ . We interpret state  $m$  as the situation in which there are  $m$  people who have not bought the article yet. In each state a seller has the actions A and N available. A recursive equilibrium point can be found in the following way:

First one analyses state 0, then state 1, and so on, up to and including M. In state 0 both players will choose N. In analysing state  $m$ , one assumes that whenever state  $m'$  ( $m' < m$ ) is reached the sellers will act in accordance with the E.P. that was found in analysing state  $m'$  (we neglect border cases in which more than one E.P. exists in some state). When one proceeds in this way, one finds the following results:

case \ state	0	1	2	3	4	5
I: $0 < c < 3/10 \cdot p$	N	A				
II: $3/10 \cdot p < c < 6/10 \cdot p$	N	N	A			
III: $6/10 \cdot p < c < 9/10 \cdot p$	N	N	N	A		
IV: $9/10 \cdot p < c < 12/10 \cdot p$	N	N	N	N	A	
V: $12/10 \cdot p < c < 15/10 \cdot p$	N	N	N	N	N	A
VI: $15/10 \cdot p < c$	N					

A indicates that both players will advertise in this situation.

N indicates that both players will not advertise in this situation.

In all cases the market is split evenly between the players.

Let us consider case V. If the sellers were able to cooperate, they certainly would not play according to this E.P. In a symmetric situation like this one it is reasonable to split the market evenly. In a cooperative situation the sellers will try to reach this in the cheapest way. Now, since everyone will

actually buy the article, advertising is useless. So the sellers will never advertise. In almost any state this cooperative behaviour is also attainable in the noncooperative situation by the use of history-dependent strategies. Namely, let the strategy pair  $(\pi, \sigma)$  be defined by:

$$\pi \left\{ \begin{array}{ll} \text{In state 0} & : \text{ do not advertise.} \\ \text{In states 5,9:} & \text{ advertise.} \\ \text{In all other} & \\ \text{states :} & \text{ if one day in the past } P^2 \text{ has advertised,} \\ & \text{ while you did not advertise that day, then} \\ & \text{ advertise; otherwise do not advertise.} \end{array} \right.$$

$\sigma$ : the same as  $\pi$ , except with  $P^2$  replaced by  $P^1$ .

It can be easily verified that the pair  $(\pi, \sigma)$  is an E.P. When the sellers act in accordance with this E.P. they cooperate implicitly. When one of the sellers breaks the agreement, he gets punished by the other. In the states 5 and 9 cooperation is not possible, since  $P^1$  cannot punish  $P^j$  heavily enough, when the latter deviates.

In the cases I, II and III the cooperative behaviour (not advertise) cannot be reached by using history-dependent strategies, since a player cannot be punished when he deviates. In these cases the only E.P. is the R.E.P.

Case IV is an intermediate case. There exists an E.P. in history-dependent strategies such that (if  $m$  is the number of the state):

- i) for small values of  $m$  the players do not advertise since the cost of advertising is too high, compared with the profits resulting from it,
- ii) for intermediate values of  $m$  the players advertise since it yields enough profit and since they cannot be punished,
- iii) for large values of  $m$  the players do not advertise since they will be punished very badly when they do not cooperate.

## 5. Linear-quadratic difference games

Formally these games do not fit into our basic model, since the state and action spaces are not finite, but are finite dimensional linear spaces. Moreover, randomization between different actions is not allowed. Furthermore, it is more convenient to work with cost, rather than with rewards. So let  $c_t^i = -r_t^i$  and  $J^i = -R^i$ .



Except for these minor points no modifications are needed to adjust the definitions of section 2 to this situation. We will not give a formal description of a Linear-Quadratic Difference Game (see e.g. [2] or [8] for a formal definition), since in the examples we will only be dealing with a special case. In example 5 we will show that we might obtain counter-intuitive solutions, when we restrict ourselves to memoryless strategies. In example 6 we will indicate that also in this situation there is a great resemblance between the cooperative and the noncooperative version of the game: all kinds of "Pareto-like" points can be reached by the use of history-dependent strategies. In this paper only deterministic games are considered. For examples of stochastic Linear-Quadratic Difference Games with better E.P.'s in history-dependent strategies we refer to van Damme ([4]).

#### Example 5

Situation:

$$x(1) \begin{matrix} u(1) \\ v(1) \end{matrix} \rightarrow x(1) + u(1) + v(1) := x(2) \begin{matrix} u(2) \\ v(2) \end{matrix} \rightarrow x(2) + u(2) + v(2) := x(3)$$

$$(x(t), u(t), v(t)) \in \mathbb{R}$$

The cost of  $P^1$  is given by:

$$J^1(x(1), u(\cdot), v(\cdot)) = (x(3) - 1)^2 + u(1)^2 + u(2)^2.$$

The cost of  $P^2$  is given by:

$$J^2(x(1), u(\cdot), v(\cdot)) = (x(3) + 1)^2 + v(1)^2 + v(2)^2$$

By dynamic programming a R.E.P. can be found. There is a unique R.E.P., which is the pair  $(\pi, \sigma)$ , with

$$\pi_1(x_1) = -\frac{2}{13}x_1 + \frac{2}{3} \quad \sigma_1(x_1) = -\frac{2}{13}x_1 - \frac{2}{3}$$

$$\pi_2(x_2) = -\frac{1}{3}x_2 + 1 \quad \sigma_2(x_2) = -\frac{1}{3}x_2 - 1$$

Now, suppose both players know that  $x(1) = 0$ . When the players play in accordance with the R.E.P., then the action taken by  $P^i$  at  $t = 1$  is  $(-\frac{2}{3})^{i+1}$  ( $i \in \{1, 2\}$ ). So each player incurs a cost of  $\frac{4}{9}$  and at time  $t = 2$  the system

is in 0 again. But, if it is known that you will be in 0 again, it is better not to take an action at all. This gives rise to the following E.P. in history-dependent strategies (given that the initial state is 0):

$$\begin{aligned}\hat{u}_1 &= 0 & \hat{v}_1 &= 0 \\ \hat{u}_2(x) &= \begin{cases} 1 & \text{if } x = 0 \\ 10 & \text{if } x \neq 0 \end{cases} & \hat{v}_2(x) &= \begin{cases} -1, & \text{if } x = 0 \\ -10 & \text{if } x \neq 0 \end{cases}\end{aligned}$$

The cost associated with this E.P. is for both players less than the cost associated with the R.E.P. Furthermore an E.P. such as the latter is intuitively more acceptable.

#### Example 6

Situation:

$$x(1) \xrightarrow[v(1)]{u(1)} x(1) + u(1) + v(1) := x(2) \xrightarrow[v(2)]{u(2)} x(2) + u(2) + v(2) := x(3)$$

$$(x(t), u(t), v(t)) \in \mathbb{R}$$

The cost for  $P^1$  is given by:

$$J^1(x(1), u(\cdot), v(\cdot)) = x(3)^2 + u(1)^2 + u(2)^2.$$

The cost for  $P^2$  is given by:

$$J^2(x(1), u(\cdot), v(\cdot)) = x(3)^2 + v(1)^2 + v(2)^2.$$

There is a unique R.E.P.  $(\pi, \sigma)$  which is given by

$$\pi_1(x_1) = -\frac{1}{3}x_1 = \sigma_1(x_1)$$

$$\pi_2(x_2) = -\frac{2}{13}x_2 = \sigma_2(x_2)$$

This R.E.P. results in cost of  $\frac{22}{169} x(1)^2$  for both players.

Let us now consider history-dependent strategies. Assume  $(\pi, \sigma)$  is an E.P. We must have that  $\pi_2(\cdot)$  and  $\sigma_2(\cdot)$  are in equilibrium for all states that

are actually reached by using  $\pi_1$  and  $\sigma_1$ . However, it is not necessary that the action pair  $(\pi_2(\cdot), \sigma_2(\cdot))$  is in equilibrium in the states that are not reached. This leads to a whole class of equilibrium points. Because of the game structure a player can punish the other as badly as he wishes. Therefore he can force the other player to steer the system to any state he wishes. So, all kinds of behaviour (even rather foolish) can appear when one plays according to a history-dependent E.P. This fact is reflected in the following class of E.P.'s:

Let  $a, b \in \mathbb{R}$  and let  $\pi$  and  $\sigma$  be defined by:

$$\pi_1(x_1) = ax_1$$

$$\pi_2(x_1, u_1, v_1, x_2) = \begin{cases} -\frac{1}{3}x_2, & \text{if } x_2 = (1+a+b)x_1 \\ \alpha|x_1| - x_2 + 1, & \text{if } x_2 \neq (1+a+b)x_1 \end{cases}$$

$$(\alpha = 9a^2 + 2(1+a+b)^2 + 1)$$

$$\sigma_1(x_1) = bx_1$$

$$\sigma_2(x_1, u_1, v_1, x_2) = \begin{cases} -\frac{1}{3}x_2, & \text{if } x_2 = (1+a+b)x_1 \\ \beta|x_1| - x_2 + 1, & \text{if } x_2 \neq (1+a+b)x_1 \end{cases}$$

$$(\beta = 9b^2 + 2(1+a+b)^2 + 1)$$

Within this class there are E.P.'s that are favourable for both players, E.P.'s that are favourable for  $P^i$  and unfavourable for  $P^j$  ( $i, j \in \{1, 2\}, i \neq j$ ) and E.P.'s that are unfavourable for both players. In a symmetric situation like this one the E.P. with  $a = b = -\frac{4}{17}$  seems a reasonable candidate for the solution of this game. This E.P. results in cost of  $\frac{2}{17}x(1)^2$  for both players. However, in a general (nonsymmetric) situation the selection of one E.P. to be the solution of the game is a difficult problem.

#### Acknowledgements

The author is grateful to ir. Jan van der Wal. The ideas behind the examples 2 and 4 are essentially due to him.

References

- [1] Aumann, R.J., "The core of a cooperative game without side payments",  
Transactions of the American Mathematical Society, 98 (1961), 539-552.
- [2] Başar, T., "On the Uniqueness of the Nash Solution in Linear - Quadratic  
Differential Games", International Journal of Game Theory, 5 (1976),  
65-90.
- [3] Bellman, R., Dynamic Programming, Princeton University Press, 1957.
- [4] v. Damme, E., "A note on Başar's: On the uniqueness of the Nash Solution",  
Memorandum Cosor M 80-06, Dept. of Mathematics, Eindhoven University  
of Technology, Eindhoven, 1980.
- [5] Groenewegen, L.P.J. and J. Wessels, "On the relation between optimality  
and saddle conservation in Markov Games" pp 183-211 in Dynamische  
Optimierung, Bonn, Math. Institut der Universität Bonn (Bonner Mathe-  
matische Schriften 98) 1977.
- [6] Kuhn, H.W., "Extensive games and the problem of information". Annals of  
Mathematics Study 28 (1953), 193-216.
- [7] Luce, D., and H. Raiffa, Games and Decisions, John Wiley and Sons, New York,  
1957.
- [8] Plasmans, J.E.J., "Linked Econometric Models as as Difference Game; Nash  
Optimality", Research Memorandum, Tilburg University, 1978.