

Infinite Histories and Steady Orbits in Repeated Games*

ITZHAK GILBOA†

*Department of Managerial Economics and Decision Sciences, J. L. Kellogg Graduate
School of Management, Northwestern University, Evanston, Illinois, 60208*

AND

DAVID SCHMEIDLER‡

*Department of Economics, Ohio State University, Columbus, Ohio, 43210; Department
of Economics and Department of Statistics, Tel Aviv University, Ramat Aviv, 69 978,
Tel Aviv, Israel*

Received May 2, 1990

We study a model of repeated games with the following features:

(a) *Infinite histories.* The game has been played since days of yore, or is so perceived by the players:

(b) *Turing machines with memory.* Since regular Turing machines coincide with bounded recall strategies (in the presence of infinite histories), we endow them with "external" memory;

(c) *Nonstrategic players.* The players ignore complicated strategic considerations and speculations about them. Instead, each player uses his/her machine to update some statistics regarding the others' behavior, and chooses a best response to observed behavior.

Relying on these assumptions, we define a solution concept for the one shot game, called steady orbit. The (closure of the) set of steady orbit payoffs strictly includes the convex hull of the Nash equilibria payoffs and is strictly included in the correlated equilibria payoffs. Assumptions (a)–(c) above are independent to a large extent. In particular, one may define steady orbits without explicitly dealing with histories or machines. © 1994 Academic Press, Inc.

* We thank Ehud Kalai, Jean-Francois Mertens, Ehud Lehrer, Dov Monderer, Abraham Neyman, Dov Samet, William Stanford, and two anonymous referees for stimulating discussions, comments, and references.

† NSF Grants IRI-8814672 and SES-9113108 are gratefully acknowledged.

‡ NSF Grants SES-9111873 is gratefully acknowledged.

The thing that hath been, it is that which shall be; and that which is done is that which shall be done: and there is no new thing under the sun.

Ecclesiastes 1:9

1. INTRODUCTION

Multiperiod decision problems in which the decision maker is faced with uncertainty are prevalent both in the scientific literature and in what is sometimes referred to as the "real world." In fact, it is difficult to give examples of "real world" problems faced by individuals or organizations which do not involve the time dimension or some uncertainty. Many models dealing with such problems may be found in the classical literature on statistical inference, dynamic programming, and repeated games. The strategies devised in these contexts tend to be very complex. Indeed, in many cases it becomes unlikely to assume that decision makers do use such strategies. Following Simon (1972, 1978), who introduced the idea that decision makers are only boundedly rational, game theorists have recently suggested models in which computational models such as finite automata and Turing machines are used to capture the intuitive notion that the decision maker can implement only strategies with a bounded complexity (defined in the appropriate sense). (See Aumann (1981), Rubinstein (1986), Neyman (1985), Ben Porath (1993), Kalai and Stanford (1985, 1988), Megiddo and Wigderson (1986), Binmore (1986), Gilboa (1988), Gilboa and Samet (1989), Abreu and Rubinstein (1988), Stanford (1987), and others.)

The paper studies three new assumptions.

a. *Infinite history.* In many cases of interest, there is no "stage 0." Organizations and individuals alike have to solve problems for which a certain history is already given. Even if an initial stage did occur somewhere in the past, the decision makers tend to *perceive* long histories as infinite ones, in the sense that they cannot recall the initial stage and reason about it. Rather, they tend to reason in such terms as: "It has always been the case that..." or "Traditionally..." and so forth. This is especially true for organizations, where a certain decision maker typically assumes his/her position long after the organization was founded. It therefore seems that an infinite-history model may be a more accurate description of reality from a psychological (or "subjective") viewpoint, even if it is less so from an actual (or "objective") perspective. (See also Schwartz (1974) for a discussion of and results on repeated games with infinite histories.)

b. *Turing machines with memory.* Modeling bounded rationality, we make one of the weakest and least objectionable assumptions, namely, that decision makers use (or can be modeled as) Turing machines. More specifically, a decision maker's reasoning is assumed to be implementable by a Turing machine (an "algorithm") which, at each point of time, starts running (from "the beginning of the program"), may consult history, and eventually halts with a decision. Note that, as opposed to finite automata, Turing machines are assumed to start their computation at each stage at their "initial state." This seems to be the natural assumption, since a "computation" or "run" of a Turing machine is a sequence of state transitions, which only gets its intended meaning when considered as a whole. Put differently, in order to benefit from the power of this computational model, one has to decouple the computation stages (i.e., the transition from state to state) from the decision stages (i.e., consecutive periods). (See the discussion in Section 5 below.) However, a very simple model presented in the sequel shows that in the context of infinite histories, a decision maker's strategy which is implementable by a Turing machine which always halts is no more than a finite recall strategy; i.e., each choice is determined by the last k periods for some $k \geq 0$. We therefore suggest strengthening the computational model by allowing some memory to be carried over from one stage to the next.

c. *Nonstrategic players.* We introduce a behavioral assumption that captures a different dimension of bounded rationality: each player considers the behavior of all other players as a Nature phenomenon, rather than strategic players. Observing a certain history, each player assumes that the others are about to play in the same way they have played in the past after similar histories. This assumption, which may also be made in a repeated game with "stage 0," is supposed to express a somewhat simplistic approach of an individual player: not knowing the other players' repeated games strategies, rather than getting involved in sophisticated reasoning about them and probability distributions on these huge spaces, the player simply assumes that they are as close to constant as possible. Of course, it would be silly on the part of the player to assume they are actually constant when he/she has contradictory evidence. However, the player chooses to believe in the simplest theory that explains his/her observations, namely, that the other players' behavior after each subhistory this player can remember is governed by a fixed distribution function. (See Gilboa (1990) for discussion and additional references.) Based on past experience, the player would compute an estimate of this distribution and choose a best response action (or mixed action) with respect to it.

Applying these three assumptions, we consider "steady orbits," defined, roughly, as follows: given a one-shot game and an integer for each

player, indicating the memory length of this player, each player is allowed to choose a (possibly different) mixed strategy (in the one-shot game) for every possible subhistory he/she may observe. Given these choices, we can compute, for every history, the probability of each play of the game in the next stage. Thus, we have a Markov chain, the states of which are (finite) subhistories.

Assuming that the game is played long enough with these strategies, each player would have the chance to estimate the distribution over the other players' moves (given his memory) quite precisely, and should his/her strategies fail to be best response, he/she would change it. A steady orbit would therefore be defined, loosely, as a selection of strategies that are a fixed point of this process.

We continue to study the set of payoffs that correspond to steady orbits. We prove that the closure of this set strictly contains the convex hull of Nash equilibria payoffs and is strictly contained in the correlated equilibria payoffs.

We note that the definition of steady orbits in repeated games is independent of the preceding analysis. Thus, one may study steady orbits by themselves, where a conceptual basis may be given by assumption (c) above and the assumption of bounded recall. (In this context, see also Lehrer (1988a,b) and Aumann and Sorin (1989).)

In Section 2 we present the basic model. This includes only one decision maker who confronts uncertainty (modeled as "nature's choice"). We also suppose that the same decision maker has lived since days of yore and will live to eternity, and that both he/she and nature choose their actions without randomization. This very primitive model captures some of the important features of a model with infinite history and these are discussed in this section.

Section 3 introduces bounded rationality into the model by defining, discussing, and studying the implications of recursive strategies. Its main point is that the standard model of Turing machine is not powerful enough to implement some intuitively simple strategies in the presence of infinite histories.

Section 4 briefly comments on the extension of the basic model to mixed moves. The definition of "Turing machine with memory" is given in Section 5. We also prove there that there are optimal Turing machines with memory for a decision maker facing a Nature phenomenon that does not "remember" more than he/she does.

In Section 6 we apply these results as a conceptual basis for the analysis of games. Relying on the implicit behavioral assumption of "nonstrategic players," we define steady orbits in repeated games and prove the results mentioned above. This section may be read separately.

Finally, Section 7 contains some concluding remarks.

2. THE BASIC MODEL FOR ONE DECISION MAKER

Let A be a finite and nonempty set of *actions* available to the decision maker (henceforth, DM) at each period. Let S denote a finite and nonempty set of possible *environments* or *states of nature* which may occur at each period. Define $C = A \times S$ to be the set of possible *circumstances*. A *history* (of circumstances) is a function $\underline{c}: \{-i | i \geq 0\} \rightarrow C$. A circumstance $\underline{c}(-i)$ will also be denoted by \underline{c}_{-i} and, when no confusion is likely to arise, by c_{-i} . Thus \underline{c} may also be written as $(\dots, c_{-i}, \dots, c_{-2}, c_{-1}, c_0)$. The set of all histories will be denoted by $C^{-\infty}$. A *future* (of circumstances) is simply an element \bar{c} of C^{∞} . We will write $\bar{c} = (\bar{c}_1, \bar{c}_2, \dots)$ or, when possible, (c_1, c_2, \dots) .

It will prove useful to define the natural projections of the set of circumstances C on A and S : let $a: C \rightarrow A$ and $s: C \rightarrow S$ be the unique functions satisfying $c = (a(c), s(c))$ for all $c \in C$. The projection functions a and s are extended to $C^{-\infty}$ and C^{∞} by the natural pointwise definition. I.e., $a(\underline{c}) = (\dots, a(c_{-1}), a(c_0))$ and $s(\underline{c}) = (\dots, s(c_{-1}), s(c_0))$ for all $\underline{c} \in C^{-\infty}$, and $a(\bar{c}) = (a(c_1), a(c_2), \dots)$ and $s(\bar{c}) = (s(c_1), s(c_2), \dots)$ for $\bar{c} \in C^{\infty}$.

We now define some operations on histories and futures:

(1) For a history $\underline{c} \in C^{-\infty}$ and $n \geq 0$, define the *n-truncation* of \underline{c} , denoted by \underline{c}^{-n*} , to be the history $(\dots, c_{-(n+1)}, c_{-n}) \in C^{-\infty}$.

(2) For a future $\bar{c} \in C^{\infty}$ and $n \geq 0$, the *n-truncation* of \bar{c} , denoted \bar{c}^{n*} , is the future $(c_{n+1}, c_{n+2}, \dots) \in C^{\infty}$.

(3) For a history $\underline{c} \in C^{-\infty}$ and $n \geq 0$, let the *n-suffix* of \underline{c} , denoted \underline{c}^{-n} , be the finite sequence $(c_{-n+1}, c_{-n+2}, \dots, c_0) \in C^n$. (For $n = 0$, \underline{c}^{-n} is always the empty string, henceforth denoted by c^0 .)

(4) For a future $\bar{c} \in C^{\infty}$ and $n \geq 0$, define the *n-prefix* of \bar{c} , denoted by \bar{c}^n , to be the finite sequence $(c_1, c_2, \dots, c_n) \in C^n$. (For $n = 0$, \bar{c}^n is always the empty string c^0 .)

(5) For a history $\underline{c} \in C^{-\infty}$ and a finite sequence $\bar{c}^n = (c_1, c_2, \dots, c_n)$, define the *concatenation* of \underline{c} and \bar{c}^n , denoted $\underline{c} \cdot \bar{c}^n$, to be the history $(\dots, c_{-1}, c_0, c_1, c_2, \dots, c_n) \in C^{-\infty}$.

Next we turn to define strategies. A *DM strategy* is a function $\sigma: C^{-\infty} \rightarrow A$. A set of all DM strategies will be denoted by Σ . *Nature's strategy* is a function $\theta: s(C^{-\infty}) \rightarrow S$. Let Θ denote the set of all nature strategies. The *future function* f maps $C^{-\infty} \times \Sigma \times \Theta$ into C^{∞} : for $(\underline{c}, \sigma, \theta) \in C^{-\infty} \times \Sigma \times \Theta$, the *future determined by* $(\underline{c}, \sigma, \theta)$, i.e., $f(\underline{c}, \sigma, \theta)$, is the element $\bar{c} \in C^{\infty}$ such that for all $n \geq 1$, $c_{n+1} = (\sigma(\underline{c} \cdot \bar{c}^n), \theta(s(\underline{c} \cdot \bar{c}^n)))$.

Given a triple $(\underline{c}, \sigma, \theta) \in C^{-\infty} \times \Sigma \times \Theta$, we say that \underline{c} is *consistent* with σ and θ if for every $n \geq 1$, $c_{-(n-1)} = (\sigma(\underline{c}^{-n}), \theta(s(\underline{c}^{-n})))$. A history \underline{c} is *possible* if there exist $(\sigma, \theta) \in \Sigma \times \Theta$ such that \underline{c} is consistent with σ and θ .

A history $\underline{c} \in C^{-\infty}$ is said to be *cyclical* if, for some $k \geq 1$, $\underline{c}^{-k*} = \underline{c}$. In this case \underline{c} will also be said to be *cyclical of order k* and \underline{c}^{-k} will be called its *cycle*. Likewise, a future $\bar{c} \in C^{\infty}$ is *cyclical* if for some $k \geq 1$ $\bar{c}^{k*} = \bar{c}$, \bar{c} is *cyclical of order k* and \bar{c}^k is its *cycle*.

We can now state two observations (the proofs of which are immediate).

OBSERVATION 2.1. *Suppose that $\underline{c} \in C^{-\infty}$ is consistent with $\sigma \in \Sigma$ and $\theta \in \Theta$, and that \underline{c} is cyclical of order k for some $k \geq 1$. Then $f(\underline{c}, \sigma, \theta)$ is also cyclical of order k and its cycle is a cyclical permutation of that of \underline{c} .*

OBSERVATION 2.2. *A history $\underline{c} \in C^{-\infty}$ is possible iff one of the following two conditions holds:*

- (i) \underline{c} is cyclical;
- (ii) For all $k \geq 0$, \underline{c}^{-k*} is not cyclical.

These observations show that our very definitions of strategies already entail some assumptions. These assumptions deserve comment. Let us first consider the definition of a DM strategy. In the bulk of literature in dynamic programming and repeated games, the decision maker's strategy is also defined as a function from histories to actions. However, in the case of finite histories, this definition constitutes no loss of generality: finite histories differ in their length. Hence, a strategy which depends only on the history also implicitly depends on "time," i.e., the stage at which the process is in. In our case, on the other hand, the same definition of a strategy does not allow the DM's choices to depend on "time," or on some "external clock." The "clock," that is, the specific enumeration of stages from $-\infty$ to ∞ is only known to the outside observer and, indeed, this enumeration may be shifted by any integer without changing the model.

This implicit "no clock" assumption may also be represented as follows: we assume that all those circumstances relevant to the DM's choice are described in A and S . That is to say, the decision maker in this model is not allowed to deviate from a certain behavior pattern "just because" time has passed. In fact, given any model with a "clock," one may construct an equivalent model without a "clock" by incorporating the state of the clock into the state of nature. If the original model's clock had infinitely many states, e.g., all integers, then S would not be finite. However, no problem would arise for finite-state clocks. (Consider, for instance, a good old-fashioned clock showing only the time of day but not the date, and so forth.)

Since our main motivation is to study some notions of "stage ∞ ," the no-clock assumption can also be explained by stating that at "stage infinity" there is no "sense of time"; put differently, $\infty + 1 = \infty$.

Let us now turn to nature's strategy. The only mathematical difference between nature and the DM is assumed to be that nature's choices do not depend on those of the DM (while the converse is, of course, false). Though philosophically questionable, this assumption seems to be a reasonable one for practical purposes. As regards the descriptive aspects of the model, it should also be noted that in many cases of interest even if this assumption fails to hold, it does describe the way the decision maker perceives the problem. For instance, weather conditions are known to be affected by people's actions; nevertheless, people tend not to take these effects into consideration while facing a decision problem. Moreover, neglecting these effects appears to give a rather good approximation to the real problem, especially if computability and complexity constraints are considered.

The fact that nature's strategy is also defined as a function of history alone (without dependence on a "clock") has a similar meaning to the corresponding definition of the DM's strategy. The assumption may seem to be restrictive, since it concerns nature for which "bounded rationality" arguments do not seem to apply (as opposed to a human being). However, this assumption is equivalent to the following: if the states of nature have been alternating in a given cycle from days of yore until today, we assume that the same pattern will persist. If no cycle has been followed throughout history, this assumption is trivially satisfied.

3. RECURSIVE STRATEGIES

We now turn to impose bounded rationality assumptions on the strategies. The main assumption we will use is that these strategies are recursive, i.e., that there exist algorithms which can compute the next action of the DM and the next state of nature given the infinite history. Before we turn to discuss these assumptions, let us first specify the computational model we use.

Our model is basically a standard Turing machine with an assignment of an action in A (or state in S) to each final state (an internal state at which the computation may terminate). We assume that the input tape always contains an infinitely-long input string. It seems more convenient to think of two tapes—the (read only) input tape and the working tape (which is always empty at the beginning of the computation). For the sake of brevity, we omit the formal definition of such a machine. However, it is a straightforward adaptation of the standard one. (See, e.g., Hopcroft and Ullman (1979).)

We also assume that the Turing machines describing the relevant strategies are such that they always halt for every conceivable input. By "conceivable" we mean that any history $\underline{c} \in C^{-\infty}$ and its projection $s(\underline{c})$ should

be considered as potential inputs to the DM's and nature's machines, respectively. It should be emphasized that it does not suffice to assume that the machines halt for every history consistent with them. For some cases, especially where the DM considers a change of strategy, we would like the future to be well-defined for any history and any pair of strategies.

Let us then denote by Σ^0 the subset of the set of all DM's strategies Σ which are implementable by a Turing machine which always halts. Let Θ^0 denote the corresponding subset of Θ .

For the DM's strategies, the restriction of allowable classes to Σ^0 seems almost innocuous, or, at least, a very weak assumption of bounded rationality. It only states that the DM's strategy can be unambiguously defined by a finite number of instructions. However, the corresponding assumption imposed on nature may seem unjustified. In fact, it is equivalent to the hypothesis that there exists a (finite) "scientific" theory which is "true" in the sense of perfectly predicting nature's choice. The philosophical grounds of such an hypothesis are beyond the scope of this paper. (See Gilboa (1990).) Nonetheless, we will not adhere to the deterministic model for very long. Once we allow for randomized choices this assumption would only mean that the *distribution over* states of nature, rather than the specific choice of one of them, is describable by a finite algorithm. It seems to us that such an assumption is not too restrictive for many cases of interest. (Consider, for instance, the concept of a Markov chain: there are several states, to each of which there corresponds a distribution over the set of states, and that distribution may be computed by an algorithm given the history.)

In the literature of repeated games there are several notions of bounded rationality. The most restrictive assumption seems to be that of *finite recall*. In our terms, $\sigma \in \Sigma$ ($\theta \in \Theta$) is said to be a *finite recall strategy* if there exists a number $n \geq 0$ and a function $f: C^n \rightarrow A$ ($f: S(C^n) \rightarrow S$) such that $\sigma(\underline{c}) = f(\underline{c}^{-n})$ ($\theta(s(\underline{c})) = f(s(\underline{c}^{-n}))$) for all $\underline{c} \in C^{-\infty}$.

A simple but somewhat surprising result is

PROPOSITION 3.1. *The set of finite recall strategies in $\Sigma(\Theta)$ is exactly $\Sigma^0(\Theta^0)$.*

Proof. The fact that every finite recall strategy is recursive is trivial. The converse is an application of Koenig's lemma and may be proved as follows: let $\sigma \in \Sigma^0$. (The proof for Θ^0 is symmetric.) Consider the infinite tree, (V, E) , which describes the conceivable histories:

$$V = \{\underline{c}^{-n} | \underline{c} \in C^{-\infty}, n \geq 0\}$$

$$E = \{(\bar{c}^{-n}, \underline{c}^{-n-1}) | \underline{c} \in C^{-\infty}, n \geq 0\}$$

That is, the vertices are all finite sequences of circumstances (c_{-n+1}, \dots, c_0) including the empty sequence c^0 which is the root of the tree, and two such histories are connected by an edge if and only if the first one may be obtained from the second by deleting the first component of the latter.

For a given $\sigma \in \Sigma^0$ and a given history $\underline{c} \in C^{-\infty}$, one may consider the path in this tree along which the computation of σ will proceed, that is, the subset of finite sequences σ may "observe" at various points of the computation. Note that all these are suffixes of \underline{c} and therefore generate a path. Since σ is required to halt, this is simply a finite path generated by \underline{c}^{-i} for $0 \leq i \leq k$ for some k . To prove that σ is a finite recall strategy, we only need to show that the length of this path, k , is bounded from above for all conceivable histories \underline{c} . Assume the contrary, i.e., that such a bound does not exist. Consider the root of the tree c^0 , and the finite number of branches emanating from it. If in each branch the computation paths of σ were bounded, the maximal bound plus 1 would have been a bound for all conceivable histories. Hence there exists at least one branch (subtree) for which there is no such bound. Continuing in this fashion one obtains a conceivable history \underline{c} for which the computation of σ does not halt. ■

OBSERVATION 3.2. Suppose $\sigma \in \Sigma^0$ and $\theta \in \Theta^0$ are consistent with the history $\underline{c} \in C^{-\infty}$. Then \underline{c} is cyclical.

Proof. In view of Result 3.1, both σ and θ are finite-recall strategies. Assume $\sigma(\underline{c})$ depends only on \underline{c}^{-k_1} and $\theta(s(\underline{c}))$ on $s(\underline{c}^{-k_2})$, and let $k = \max\{k_1, k_2\}$. It is trivial to see that \underline{c} is cyclical of order $m \leq |C|^k$. ■

The last two results show that our model is too restrictive, since it can only describe cyclical phenomena. We are now going to generalize it in two ways by introducing (i) machines with memory and (ii) randomized actions. However, the main two assumptions, that of time stationarity and that of bounded rationality, will essentially be retained.

4. RANDOMIZED ACTIONS

As mentioned in Section 3, the assumption that nature has a recursive strategy (with the "no clock" assumption) seems far too restrictive to describe uncertainty. On the other hand, we would not like to drop this assumption altogether since we are interested in modeling situations in which there is something the decision maker may infer from the past regarding the future.

The most natural thing to do at this point seems to be to allow randomized actions (at least for nature), and to require that the assumptions discussed above be satisfied with respect to these, rather than the actual

actions (or states of the world) chosen. Thus, no conceivable history will be ruled out and, in particular, the model will not be restricted to cyclical phenomena—but there will still be enough regularity for the DM to apply statistical inference techniques.

Let $\Delta(A)$ and $\Delta(S)$ denote the set of distributions over A and S , respectively, and define $\Sigma^* = \{\sigma: C^{-\infty} \rightarrow \Delta(A)\}$; $\Theta^* = \{\theta: s(C^{-\infty}) \rightarrow \Delta(S)\}$. As in the deterministic model, these definitions already impose the “time stationarity” or “no clock” assumptions. In order to formulate the bounded rationality assumption we will define Turing machines as before, save that now a distribution over A (or S) will be attached to each final internal state of the machine rather than a single element of it. Note that these distributions may involve irrational numbers, since these probabilities are not assumed to be computed by the machine. This modeling assumption is consistent with our general application of the Turing machine model: we only use it to bound the *strategic* complexity of a strategy. The complexity of tossing a “coin” with probability $\sqrt{2}/2$ of “head” seems to us to be a technical, rather than a substantial difficulty. In other words, the idealized Turing machine we use tries to capture faithfully the way people make decisions. While people cannot follow strategic arguments beyond a certain complexity level, we do not seem to be befuddled by a $(\sqrt{2}/2; 1 - \sqrt{2}/2)$ randomization, as opposed to the (.7, .3) one. Obviously, actual decisions will not typically follow irrational probabilities. But imposing their computation on the same machine would confound two types of complexity, only one of which seems to be of significance.

(Alternatively, one may assume that each terminal state of our machine sets into motion a separate, auxiliary machine, whose role is to compute an action with approximated probabilities. This is tantamount to having one “large” machine with two types of states, only one of which is counted in complexity measures such as the number of states and length of computation.)

5. MACHINES WITH MEMORY

Result 3.1 may suggest that recursive strategies are too restrictive. In fact, they cannot implement reasonably simple strategies. Consider the following example: nature chooses whether it will rain or not. The DM chooses to take an umbrella or not. A possible DM strategy is the following “2-trigger strategy”: if it never rained before, or if it rained exactly once in the entire history, do not take an umbrella. Otherwise, take an umbrella. This strategy is certainly not a finite recall strategy. Hence, it is not recursive. However, such a strategy is implementable by a finite automa-

ton playing a repeated game as in Aumann (1981), Rubinstein (1986), Neyman (1985), Kalai and Stanford (1985), and others. (These models assume that histories are finite, but the extension of the automaton notion to infinite histories is straightforward.)

This apparently paradoxical result, namely that a Turing machine is weaker than a finite automaton, is due to an abuse of terms: when a finite automaton is said to "implement" a strategy in a repeated game, it is meant that each stage of the game corresponds to one application of the automaton's transition function. When a Turing machine is said to "compute" a strategy, the role of the machine's internal states is quite different: they are used for a "background" computation, and only when the whole computation is completed is an action chosen and one single stage of the game is over. In other words, the automaton may use its states to "remember" information from one stage to the next. The Turing machine uses its states for the computation alone, and at each stage it is assumed to begin at the same state. Thus it cannot carry information bits computed in previous stages to the next one. In the context of finite histories, this lack of memory on the Turing machine's part constitutes no loss of generality: the machine can always simulate its computations in the previous stages. (This is, of course, only true if computability rather than complexity aspects are taken into account.) However, in the case of infinite histories this is no longer the case. It therefore seems natural to consider a larger set of machines which have another "memory" tape which may be written to and read from during a computation. This tape is kept unaltered between the end of one computation and the beginning of the next.

We would therefore like to endow a DM's machine with memory, and it would seem suitable to allow it to remember real numbers such as relative frequencies. However, if we do not impose any additional restrictions and consider a Turing machine with real-valued registers on which arithmetic operations and comparisons can be performed, we make it significantly stronger than we originally suggested: such a machine can use a real number as a code and analyze it to determine its behavior. Thus, a very simple machine may actually implement a nonrecursive strategy. This is obviously not what we had in mind. However, if we limit the memory to a finite number of cells (each of which may contain one of a finite set of symbols) we may be unduly restricting the machine's ability to compute numbers in a naive way.

We therefore have to distinguish between the memory the machine has for numbers per se and numbers as encoding of information. We were unable to find an elegant computational model which will draw this distinction by its computational abilities. Rather, we suggest adopting a Turing machine with several real-valued registers, but restricting its complexity. That is, each computation performed by the machine may use the registers

up to T times for some fixed T (which will be a part of the machine's specifications).

We therefore define an m - T -Turing machine (for $m, T \geq 0$) as a machine with m real-valued registers such that each computation may involve no more than T register-accesses, each of which may be an arithmetical operation ($+$, $-$, \times , $/$) or a comparison ($=$, \neq , $>$, \geq , $<$, \leq).

More specifically, the transition function of such a machine may depend not only on the "current" state and the literal read on the tape, but also on an arithmetical comparison of the type: "Is the number in register 5 greater than that in register 17?" Similarly, the transition function may specify, on top of the next state, the literal to be written on the tape and the "head" movement—also an arithmetic operation such as "add the content of register 2 to that of register 5 and put the result in register 11." An instantaneous description of such a machine includes the (real-valued) content of all its registers, and a "computation" or a "run" is a sequence of instantaneous descriptions, where each one of these follows from its predecessor according to the intended meaning of the arithmetic operations and comparisons. The formal definition of such a machine, a computation of it, and so forth, are omitted. They are straightforward adaptations of the classical definitions. (See also Preparata and Shamos (1985), pp. 28–29.)

A few words on this computational model may be in place here. As in Section 4 above, we make here an assumption of an idealized Turing machine which has various "black boxes." These black boxes perform the most astonishing computational tasks in no time, and one may obviously wonder how they "really" are done.

But this question is close to irrelevant for our purposes. The metaphor of a Turing machine should not obscure the modeling target. It is not an actual computational device, nor the theoretical construct that we attempt to study. The Turing machines used in our model are supposed to be abstracted, simplified representations of human decision makers. They should be judged on the basis of their psychological plausibility rather than actual accuracy. Thus, we employ the machine metaphor when it seems to be an appropriate measure of strategic complexity, but discard it when it introduces other complexity considerations which we find of little relevance. In short, we are not committed to Turing machines or any other modeling tool. We simply attempt to find a model that would best reflect important features of reality, while neglecting the less important ones.

We may now turn to some relevant properties of m - T -Turing machines.

OBSERVATION 5.1. *Suppose M is a m - T -Turing machine which computes an action $a \in A$ (or $a \in \Delta(A)$) for every possible history and for*

every set of initial values in its registers. Then there exists a constant $k(M)$ such that for every history $\underline{c} \in C^{-\infty}$, M does not consult $\underline{c}^{-k(M)*}$.

Proof. Similar to that of Proposition 3.1. ■

Therefore, we can conclude again that, apart from the memory the machine carries, it is still a bounded recall machine. However, such machines can already learn a stochastic process. That is, suppose that Nature's strategy is a finite recall mixed strategy $\theta \in \Theta^*$. Assume that θ depends only on the last k stages, and that $h: A \times S \rightarrow \mathbb{R}$ is the DM's payoff function. (Notice that we do not allow nature's strategy to have memory as well. This should be interpreted as a bounded rationality assumption: the DM will be optimizing only against "simple" phenomena.) For every history \underline{c} we consider

$$\sup_{\sigma \in \Sigma^*} \liminf_{T \rightarrow \infty} 1/T \sum_{t=1}^T E(h(f_t(\underline{c}, \sigma, \theta))),$$

where f is the future function defined in Section 2 above.

A strategy σ is (ε -) optimal against θ if it (ε -) obtains this payoff for all \underline{c} . An m - T -Turing machine is (ε -) optimal against θ if for every set of initial values assigned to its registers $\{x_i\}_{i=1}^m$, the induced strategy $\sigma_{\{x_i\}}$ is (ε -) optimal against θ .

The fact that payoff measurement "begins" at a given (infinite) history \underline{c} poses a problem since it may implicitly introduce a "clock": one may design a machine which uses a certain register as a "timer": if it is positive, it is increased by 1, and otherwise it is set to 1. Thus, for every history and every set of register-values $\{x_i\}_{i=1}^m$, such a machine may practically decide to start "counting" at that point. We find this to contradict our notion of an infinite history, since the point of time at which we decide to start the computation of the payoff is arbitrary, and the machine's computation should not depend on it.

We therefore define an m - T -Turing machine M_0 to be θ -consistent (for $\theta \in \Theta$) if there is an infinite history \underline{c} and a history of register values $\{x_i^{-t}\}_{i=1}^m\}_{t \geq 1}$ such that for every $t \geq 1$, beginning with history \underline{c}^{-t*} and the register values $\{x_i^{-t}\}_{i=1}^m$, M_0 and θ could generate the continuation \underline{c}^{-t} and the register values with positive probability.

Notice that the type of machines described above is not θ -consistent for any $\theta \in \Theta^*$: the history of the counter registers should be of the form $(\dots, 0, 0, 0, 1, 2, 3, \dots)$ which is incompatible with the definition of θ -consistency. Intuitively, a (θ -)consistent machine cannot "decide" to start counting at an arbitrary point. Put differently, if it never started counting up to a given point, it will never start thereafter as well. (In the example above, it would never generate a sequence of zeros.)

The following two results guarantee the existence of ε -optimal and optimal (θ -)consistent machines against “simple” nature phenomena, that is, strategies θ with finite recall. We will show that, for every given recall length, there is a machine which is (ε -)optimal against *any* phenomenon whose recall does not exceed the given bound.

Both results are standard in case of a finite history, and are based on statistical estimation of θ . ε -optimality may be obtained by a long enough sample, and optimality by an ever-increasing sample size. However, infinite histories pose two problems: (i) How does one update statistics gathered over infinitely many periods? (ii) How can a (θ -)consistent machine which has no “clock” increase its sample size?

The answers are given in the following results.

LEMMA 5.2. *Given nonempty and finite sets A , S , a payoff function $h: A \times S \rightarrow \mathbb{R}$, an integer $k \geq 0$ and $\varepsilon > 0$, there exists an m -T-Turing machine M with $k(M) = k$ which is θ -consistent for and ε -optimal against every $\theta \in \Theta^*$ with recall k .*

Remark. Notice that the DM implicitly assumes that nature does not use more recall than he/she does. Again, this is a bounded rationality assumption: knowing very little on what nature “really” is, the DM does his/her best.

Proof. As explained above, all M needs to do is to compute the relative frequency of each state of Nature s after each k -suffix of a history $\underline{s} \in (S)^k$. For each \underline{s} that θ may generate with a positive probability, the relative frequency of each $s \in S$ will approximate its true (θ) probability of following \underline{s} with a high probability. Thus, $|S|^{k+1}$ registers are all that is needed for ε -optimality. The problem is how to compute the relative frequencies given an infinite history and one additional observation. The idea of the solution is the following: the machine will have to keep and respond to statistics gathered over finitely many periods, where their minimal number is determined by the desired level of (ε -)optimality. In order to guarantee that at each point of time the machine has a long enough history, it would actually have two sets of registers. Each one would update statistics for many periods, then “forget” its data and start anew (from scratch). However, if the two cycles are appropriately synchronized, when one of registers “forgets” all it knows, the other still has enough memory and will also keep it for a while. Thus, the machine may play ε -optimally at any given point of time, yet “learn” any nature strategy θ .

Formally, let $\varepsilon_1 > 0$ be given, and let $N \geq |S|^2/\varepsilon_1^3$. It follows from Markov’s inequality that if X_1, X_2, \dots, X_n are i.i.d. random variables, each of which is distributed over S , and $\mu \in \Delta(S)$ is their expectation, then

$$P\left(\left\|\left(1/n\right)\sum_{i=1}^n X_i - \mu\right\| \geq \varepsilon_1\right) \leq \varepsilon_1, \text{ for all } n \geq N,$$

(where $\|\cdot\|$ is the Euclidean norm). Construct a machine M with $2 \times (|S|^{k+1} + |S|^k) + 10$ registers as follows. At any point of time, for each history suffix \underline{s} of length k and for each state of nature s_0 , the machine would have two relative frequencies for the occurrence of s_0 after \underline{s} : one obtained over the last N_0 occurrences of \underline{s} and the other over $N_0 + N$ last occurrences, for some N_0 with $0 \leq N_0 \leq N - 1$. (Thus, for each \underline{s} and in each of the two sets of registers it has $|S|$ registers for the distribution over S and one for the counter.) At each stage the machine uses the statistics obtained over the longer history for choosing a best response action, and updates both relative frequencies of the corresponding $\underline{s} \in (S)^k$. Correspondingly, it advances the counters of that \underline{s} , say N_0 and $N_0 + N$, by 1. When $N_0 + N = 2N$, the machine sets N_0 to zero and ignores the relative frequencies obtained over the longer history. (That is, it clears the corresponding registers.) The ten additional registers are used to retain the constants N , 1, 0, as well as for auxiliary computations.

Thus, at each stage the machine has a relative frequency which is ε_1 -close to the true Nature strategy, and it chooses a best response act versus the approximated distribution.

Obviously, for small enough ε_1 , i.e., for large enough N , the machine obtains an ε -optimal payoff. Finally, note that, due to its cyclical nature, this machine is also θ -consistent. ■

We now strengthen the previous lemma by “dropping the ε .”

THEOREM 5.3. *Given nonempty and finite A , S , a payoff function $h: A \times S \rightarrow \mathbb{R}$ and an integer $k \geq 0$, there exists an m -T-Turing machine M with $k(M) = k$ which is θ -consistent for and optimal against every θ with recall k .*

Proof. We first observe that the recall length and the number of registers of the machine constructed in 5.2 above is independent of ε . That is, ε determines the length of sample size N , but the same m registers will be used the same number ($\leq T$) of times for all $\varepsilon > 0$. This fact raises the obvious solution of increasing the sample size: N is stored in one of these registers which, whenever a cycle is completed (i.e., when $N_0 = N$), is incremented by 1 (or, say, multiplied by 2).

Thus, the machine would play ε_n -optimally for a sequence $\varepsilon_n \rightarrow 0$, and (by the liminf criterion) also optimally.

But this suggested construction does not seem to satisfy the consistency condition: if the machine increases N , its history will contain some “starting point” at which N is zero. How would the machine have known it was then time to start counting?

The solution is quite straightforward. Let the counter N have any integer value, including negative values. The sample size the machine actually uses will be $|N|$, and whenever $N_0 = |N|$, N is increased by 1. Thus, the precision of the statistics is increasing in both time directions. Intuitively speaking, such a machine learned θ precisely in the past, has since forgotten it, and is in the process of learning it again. Obviously, for every given history and every set of register-values, N will eventually become positive and the machine's play will be closer and closer to optimality. With the \liminf criterion, optimality is guaranteed.

Formally, construct a machine that is identical to that described in 5.2, with an additional register for the value of $|N|$. This is the register to which N_0 is compared during the cycle. If $N_0 = |N|$, N (which may be negative) is incremented by 1, and $|N|$ is computed again. The rest of the algorithm is unaltered.

Notice that this machine will be θ -consistent for every θ . ■

6. STEADY ORBITS IN GAMES

Consider a one-shot finite normal form game $G = (N, (S^i)_{i \in N}, (h^i)_{i \in N})$ where $N = \{1, 2, \dots, n\}$ ($n \geq 1$) is the set of *players*, S^i is a (finite and nonempty) set of *moves* of player i , and $h^i: S \rightarrow \mathbb{R}$ is player i 's *payoff* function where $S = \prod_{i \in N} S^i$.

We would like to define a solution concept for one-shot games which would rely on the implicit assumptions that G is an infinitely repeated game with infinite history and that each player considers all other players (in conjunction) as Nature. For brevity's sake we will not provide formal definitions of the repeated game and its strategies. These definitions are straightforward adaptations of the ones given above. However, bearing this interpretation in mind and applying the previous sections' results we know that a recursive strategy with a bounded number of memory registers is, in fact, a bounded recall strategy (with the same number of memory registers). Let $k_i \geq 0$ denote player i 's recall, i.e., the number of periods player i remembers. Thus our solution concept will be defined for a given one-shot game G and a given recall profile $\{k_i\}_{i \in N}$.

By the previous section's results we know that, if a player i believes he/she faces a nature phenomenon with recall k_i , he/she has an optimal "statistical" strategy (implementable by an m - T -Turing machine). We therefore assume that each player will choose such a strategy, as if he/she were implicitly assuming—for lack of better workable hypothesis—that nature does not use more than k_i recall stages. Of course, m - T -machines may implement a variety of other bounded-recall strategies. However, we assume that the players are "rational" enough to optimize given their

beliefs. (As in most game-theoretic models, the rationality assumption and the players' choices will not be formally modeled. Rather, they will be implicit in the definition of the solution concept.)

Our focus will be on the limit frequency of each $s \in S$ during a play of the game by such players as well as on the corresponding average payoffs.

We first define the set of player i 's (mixed) strategies to be

$$\Sigma^i = \{\sigma^i: (S)^{k_i} \rightarrow \Delta(S^i)\}.$$

Given the recall bounds $\{k_i\}_{i \in N}$ we define $k = \max_i k_i$. We will be interested in states which are the $|S|^k$ k -tuples of one-shot plays; those summarize the relevant information of the history.

Some notational conventions which will prove useful are the following. For a set X (such as S^i , S , etc.) and $\underline{x} \in X^m$, say $\underline{x} = (x_1, x_2, \dots, x_m)$, we define $\text{suf}(\underline{x}, l)$ ($1 \leq l \leq m$) as the element of X^l defined by $(x_{m-l+1}, x_{m-l+2}, \dots, x_m)$. Likewise, $\text{pref}(\underline{x}, l)$ will denote $(x_1, x_2, \dots, x_l) \in X^l$. For $\underline{x} = (x_1, \dots, x_m) \in X^m$ and $\underline{y} = (y_1, \dots, y_l) \in X^l$ we define $\underline{x} \cdot \underline{y}$ to be the element $(x_1, \dots, x_m, y_1, \dots, y_l) \in X^{m+l}$.

We also define $\Sigma = \prod_{i \in N} \Sigma^i$ and a typical element $\sigma \in \Sigma$ will be understood to define $\sigma^i \in \Sigma^i$ (for all i) such that $\sigma = (\sigma^1, \sigma^2, \dots, \sigma^n)$. For $i \in N$ we define $s^{-i}(\sigma^{-i})$ to be an element of $S^{-i} \equiv \prod_{j \neq i} S^j$ ($\Sigma^{-i} \equiv \prod_{j \neq i} \Sigma^j$), the symbol $(s^{-i})^j ((\sigma^{-i})^j)$ will stand for player j 's component in $s^{-i}(\sigma^{-i})$. The symbols (s^{-i}, t^i) and (σ^{-i}, τ^i) would be elements of S and Σ , respectively, with the obvious meaning.

Given $\sigma \in \Sigma$ we define a Markov chain whose set of states is $(S)^k$ with the transition probability matrix $A(\sigma)$ defined by the elements:

$$a(\sigma)_{s, s'} = \begin{cases} \prod_{i \in N} \sigma^i(\text{suf}(\underline{s}, k_i))((s'_i)^i), & \text{if } \text{suf}(\underline{s}, k-1) = \text{pref}(\underline{s}', k-1) \\ 0, & \text{otherwise.} \end{cases}$$

In other words, $a(\sigma)_{s, s'}$ is the conditional probability that the play of the game will have a history (of length k) \underline{s}' at time t , given that at time $(t-1)$ it has a history \underline{s} and the players are playing according to σ .

The probabilities in $A(\sigma)$ describe all the relevant information about the play of the game given a certain history. However, as explained below, we also have to introduce the stationary distribution as a part of the definition of a steady orbit.

Let there be given $p \in \Delta((S)^k)$ and $\sigma \in \Sigma$. We define a function $\tau^i(p, \sigma): (S)^{k_i} \rightarrow \Delta(S^{-i})$, which should be interpreted as the list of values of the statistics player i retains in his/her machine registers. More specifically, should the players choose σ , and if the game has been played according to the stationary distribution p on $(S)^k$, for a given $\underline{s} \in (S)^{k_i}$ $\tau^i(p, \sigma)(\underline{s})$ is a distribution on S^{-i} . For each $(n-1)$ -tuple of the other

players' moves $s_0^{-i} \in S^{-i}$, $\tau^i(p, \sigma)(\underline{s})(s_0^{-i})$ is the relative frequency with which s_0^{-i} followed \underline{s} . Given player i 's recall, k_i , the distribution $\tau^i(p, \sigma)(\underline{s})$ reflects player i 's belief about what the environment will be in the next stage, and this is the distribution i will be choosing a best response to.

In case all players have identical recall ($k_i = k$), $\tau^i(p, \sigma)(\underline{s})$ will also be a correct prediction, namely, the actual distribution on the next $(n - 1)$ -tuple of the other players' moves. In particular, it is determined by what the players play—namely, σ —and the current state \underline{s} , and independent of the stationary distribution p . More specifically, if $k_i = k$,

$$\tau^i(p, \sigma)(\underline{s})(s_0^{-i}) = \prod_{j \neq i} \sigma^j(\underline{s})(s_0^{-i})^j;$$

that is, the probability of the next $(n - 1)$ -tuple of moves to be s_0^{-i} is simply the product of the probabilities that each of the other players will play his/her part.

However, in case player i has $k_i < k$, he/she can only observe the last k_i states of a k -long history. Correspondingly, the statistics player i 's machine retains can only be for $(S)^{k_i}$. Given $\underline{s} \in (S)^k$, however, there are $(|S|)^{k-k_i}$ k -long histories $\underline{s}' \in (S)^k$ which give rise to it. Therefore, the distribution $\tau^i(p, \sigma)(\underline{s})$ is a weighted average of the actual distributions corresponding to the relevant k -long histories \underline{s}' , where the stationary distribution p determines the weights. The general definition of $\tau^i(p, \sigma)$ is, therefore

$$\tau^i(p, \sigma)(\underline{s})(s_0^{-i}) = \sum_{\underline{s}' \in (S)^{k-k_i}} \bar{p}(\underline{s}') \prod_{j \neq i} \sigma^j(\text{suf}(\underline{s}' \cdot \underline{s}, k_j))(s_0^{-i})^j$$

where

$$\bar{p}(\underline{s}') = p(\underline{s}' \cdot \underline{s}) / \sum_{\underline{t}' \in (S)^{k-k_i}} p(\underline{t}' \cdot \underline{s})$$

if the denominator does not vanish. In case it does, i.e., $p(\underline{t}' \cdot \underline{s}) = 0$ for all $\underline{t}' \in (S)^{k-k_i}$, the definition of τ^i is immaterial.

Given a distribution on the other players' moves $q^{-i} \in \Delta(S^{-i})$, and a mixed move for player i , $q^i \in \Delta(S^i)$, we say that q^i is a *best response* to q^{-i} if it maximizes $E(h^i)$ given q^{-i} .

At last we can define a *steady orbit* of the game G with recall profile $\{k_i\}_i$ to be a pair (p, σ) such that:

- (i) $[A(\sigma)]'p = p$ (where “ $'$ ” indicates transpose);
- (ii) $\forall i \in N, \forall \underline{s} \in (S)^k, \sigma^i(\underline{s})$ is a best response to $\tau^i(p, \sigma)(\underline{s})$.

The first condition states that p is indeed a stationary distribution for the Markov chain defined by $A(\sigma)$. The second condition simply requires that all players will choose a best response (mixed) move, given their information, while their information is consistent with the stationary distribution p and all players' strategies.

It is important to note that the interpretation we suggest for this concept does not assume that the players are aware of the stationary distribution p , or even of the other players' recall bounds $\{k_i\}_i$, let alone other players' strategies. Each player gathers information and computes relative frequencies to the best of his/her ability, without knowing whether his/her bounded recall is large enough or not. If the players choose a certain strategy σ , an outside observer could define the resulting Markov chain and compute the stationary distribution. Should such an observer compute $\tau^i(p, \sigma)$ as defined above, (s)he would find that it coincides with the statistics computed by player i , but this computation of $\{\tau^i\}$ via p and σ cannot be done by the players themselves.

The first question which arises at this point seems to be existence. And indeed, a rather standard fixed-point argument ensures that the following holds:

THEOREM 6.1. *Given a one-shot game $G = (N, (S^i)_{i \in N}, (h^i)_{i \in N})$ and $k_i \geq 0$ for $i \in N$, G has a steady orbit (p, σ) with recall profile $\{k_i\}_{i \in N}$.*

Proof. Let $B = \Delta((S)^k) \times \prod_i \Delta(S^i)^{|S|^k}$. Note that each element $b \in B$ can be interpreted as a pair (p, σ) where $p \in \Delta((S)^k)$ and σ is an n -tuple of strategies. Furthermore, B is a convex and compact subset of \mathbb{R}^m for some $m (= |S|^k + \sum_{i=1}^n |S|^k |S|^i)$. Let $f: B \rightarrow 2^B$ be the following correspondence: for $(p, \sigma) \in B$, $f((p, \sigma))$ is the set of all pairs (p', σ') such that: (i) $p' = [A(\sigma)]'p$, and (ii) $(\sigma')^i$ is a best response strategy to $\tau^i(p, \sigma)$. (That is, the first component of all points in $f(p, \sigma)$ is the same p' .) Note that f is convex valued and upper semicontinuous. Hence, it has a fixed point by Kakutani's theorem. ■

Given a stationary distribution $p \in \Delta((S)^k)$ we define \hat{p} to be the induced distribution on S , i.e., $\hat{p}(s) = \sum_{s' \in (S)^{k-1}} p(s' \cdot s)$. We will also define $h^i(\hat{p})$, for $i \in N$, to be $\sum_{s \in S} \hat{p}(s) h^i(s)$ and $h(\hat{p})$ will denote the vector of expected payoffs $(h^i(\hat{p}))_{i \in N}$. With a convenient abuse of notation, we will also use $h^i(p)$ and $h(p)$ for $p \in \Delta((S)^k)$ referring to $h^i(\hat{p})$ and $h(\hat{p})$, respectively.

We will be interested in

$$SO(k_1, \dots, k_n) = \left\{ h(\hat{p}) \mid \begin{array}{l} (p, \sigma) \text{ is a steady orbit of } G \text{ with recall} \\ \text{profile } \{k_i\}_{i \in N} \text{ for some } \sigma \in \Sigma \end{array} \right\}$$

and

$$SO = \cup_{(k_1, \dots, k_n) \in (Z^+)^n} SO(k_1, \dots, k_n).$$

with $Z^+ = \{0, 1, 2, \dots\}$.

Let us denote by NE (CE) the set of Nash (correlated) equilibria payoffs (see Nash (1951) and Aumann (1974)). We will now state and prove some results regarding the relationships between the set of steady orbit payoffs and Nash/correlated equilibria payoffs. All these results hold for any given one-shot game $G = (N, (S^i)_{i \in N}, (h^i)_{i \in N})$.

PROPOSITION 6.2. *For all k_1, \dots, k_n ,*

$$NE \subseteq SO(k_1, \dots, k_n).$$

Furthermore, $NE = SO(0, \dots, 0)$, hence $NE = \cap_{(k_1, \dots, k_n) \in (Z^+)^n} SO(k_1, \dots, k_n)$.

Proof. Given k_1, \dots, k_n and a Nash equilibrium of G , define σ to be the n -tuple of strategies which play the given equilibrium regardless of the history. Let p be some stationary distribution of $A(\sigma)$ and note that (p, σ) is a SO of G with $\{k_i\}_{i \in N}$.

For the "furthermore" part, let $k_i = 0$ for $i \in N$, and assume (p, σ) is a steady orbit of G with (k_1, \dots, k_n) . Note that σ^i is no more than a mixed move in G (since there is only one zero-length history) and correspondingly, the Markov chain has only one state. The best response condition implies that σ induces a NE in G . ■

PROPOSITION 6.3. *$Co(NE) \subseteq \overline{SO}$. (Co stands for convex hull, and the bar for closure in the standard topology.)*

Proof. Let there be given m Nash equilibria of G denoted $\sigma NE_j = (r_j^i)_{i=1}^N$ ($i \leq j \leq m$), where $r_j^i \in \Delta(S^i)$. We will denote by NE_j the payoff vector $h(\Pi_i r_j^i)$ (where $\Pi_i r_j^i$ is the product distribution on S defined by the distributions r_j^i on S^i). It suffices to show that all rational convex combinations of $\{NE_j\}_{j=1}^m$ are in \overline{SO} . Let there be given, then, positive integers $\{t_j\}_{j=1}^m$ with $T = \sum_{j=1}^m t_j$. We would like to show that $\sum_{j=1}^m (t_j/T) NE_j \in \overline{SO}$.

Let us first explain the main idea of the proof. The obvious way to obtain the desired average payoff is to let all players have identical recall k and play one of the Nash equilibria each stage. Since all players would observe the same history, each one will know the exact mixed move of every other player and the strategies will be best-response ones.

Let us first consider the case in which all the Nash equilibria are pure. In this case it suffices to set $k = T$ and let the players play

σNE_j t_j times ($j = 1, \dots, m$) in a cycle. However, the same technique cannot be directly applied to mixed equilibria since the history (i.e., the actual realizations) does not identify the Nash equilibria that were played, and one cannot uniquely define the next equilibrium which should now be played.

We note here that the difficulty could be avoided if one assumes that each player remembers his/her own mixed actions. In this case the proof is identical to the pure Nash equilibria case. However, one need not deviate from our framework in order to obtain the result, and we therefore stick to it.

The main idea, which is not very surprising, will be the following: instead of a single play of an equilibrium (which does not identify it) we should have a long sequence of plays, which will identify it with high enough probability.

Thus, a history in which each equilibrium was played long enough according to a certain cycle is likely to regenerate a similar history. We should also verify that histories that do not correspond to the desired cycle will not have too high a (stationary distribution) probability; we can guarantee that by deciding that the players would play an arbitrarily chosen σNE , say σNE_1 , at those histories.

The formal proof is, naturally, slightly more delicate: let there be given $\{\sigma NE_j\}_{j=1}^m$ (without loss of generality, $NE_i \neq NE_j$ for $i \neq j$), $\{t_j\}_{j=1}^m$ and $1 > \varepsilon_0 > 0$, with $T = \sum_{j=1}^m t_j$.

Let us define

$$d = \min_{i \leq i \neq j \leq m} \|\sigma NE_i - \sigma NE_j\| > 0,$$

(where $\|\cdot\|$ stands for the Euclidean norm).

Without loss of generality we will assume that for every $s \in S$, $\|h(s)\| < \frac{1}{2}$, hence also for every $s, s' \in S$ $\|h(s) - h(s')\| \leq 1$. For $x \in \Delta(S)$ and $\varepsilon > 0$ denote $N_\varepsilon(x) = \{y \in \Delta(S) \mid \|x - y\| < \varepsilon\}$.

Let M be an integer satisfying $M > 3/d$. For such an M , if $x_1, x_2, \dots, x_M \in \Delta(S)$ and $h(x_1), h(x_2), \dots, h(x_{M-1}) \in N_{d/3}(\sigma NE_j)$, then

$$1/M \sum_{i=1}^M h(x_i) \notin N_{d/3}(\sigma NE_i) \quad \text{for } i \neq j.$$

Next, let $\varepsilon_1 = \varepsilon_0/3m(m-1)$ and choose L to be an integer such that $L > 1/\varepsilon_1$.

Let K_0 be a large enough integer such that for all $k \geq K_0$ and every $1 \leq j \leq m$, if σNE_j is played k times, forming a sequence X_1, X_2, \dots, X_k of i.i.d. random variables on $h(\Delta(S))$, then

$$\text{Prob}(\|(1/k) \sum_{i=1}^k X_i - NE_j\| < d/3) > (1 - \varepsilon_j)^{1/ML(T+1)}.$$

A “long sequence” of a certain equilibrium will be of length $K_0 ML$, and thus we define the recall length to be $k = K_0 ML(T + 1)$. For each player $i \in N$, then, $k_i = k$.

We need some additional definitions: a sequence $\underline{s} \in (S)^{MK_0}$ is an M - j -sequence if

$$(1/K_0) \sum_{r=1}^{K_0} h(s_{qK_0+r}) \in N_{d/3}(NE_j), \quad \text{for all } q = 0, 1, \dots, M-1.$$

(Here and in the sequel we do not distinguish between an element of S and its corresponding element of $\Delta(S)$.)

It is helpful to note at this point that

CLAIM. If $\underline{s} \in (S)^{2MK_0}$ is a concatenation of two M - j -sequences, it cannot contain a subsequence which is an M - i -sequence for $i \neq j$.

Proof of Claim. To see this, suppose that (s_1, \dots, s_{MK_0}) and $(s_{MK_0+1}, \dots, s_{2MK_0})$ are M - j -sequences and that $(s_r, \dots, s_{r+MK_0-1})$ is an M - i -sequence for some $1 \leq r \leq MK_0$. Let $0 \leq l < M$ be such that $lK_0 \leq r < (l+1)K_0$. Then:

$$(1/MK_0) \sum_{t=r}^{r+MK_0-1} h(s_t) = (1/M) \left[(1/K_0) \left(\sum_{t=r}^{(l+1)K_0-1} h(s_t) + \sum_{t=(l+M)K_0}^{r+MK_0-1} h(s_t) \right) + \sum_{v=1}^{M-1} (1/K_0) \sum_{t=(l+v)K_0}^{(l+v+1)K_0-1} h(s_t) \right].$$

By the definition of M - j -sequence, the last $(M-1)$ terms in the outer summation (in square brackets) are in $N_{d/3}(NE_j)$. The first term being somewhere in $\Delta(h(S))$, the average of them all cannot be in $N_{d/3}(NE_i)$, in contradiction to the assumption that $(s_r, \dots, s_{r+MK_0-1})$ is an M - i -sequence. ■

We can now proceed to define longer sequences. A sequence $\underline{s} \in (S)^{MLK_0}$ is an L - M - j -sequence if it is the concatenation of L M - j -sequences. We can finally turn to define a “nice” history, in which the desired Nash equilibria seem to have been played according to the correct order. Basically, it is a sequence of T L - M - j -sequences (with the j ’s taking the value i exactly t_i times), but we have to take into account the following points: (i) the sequences of different equilibria may have some overlap (our choice of M and the claim above would guarantee that this overlap is no longer

than MK_0 , i.e., $1/L$ of an L - M - j -sequence); (ii) the end (most recent part) of the sequence does not have to belong to any L - M - j -sequence (but rather to a new one in the process of formation); (iii) the order of the Nash equilibria in the sequence is

L - M - l -sequence, \dots , L - M - l -sequence, \dots ,
 t_l times \dots
 L - M - m -sequence, \dots , L - M - m -sequence
 t_m times

in some cyclical permutation.

We therefore define, for $0 \leq c \leq T - 1$, a history, $\underline{s} \in (S)^k$ to be a k - c -sequence if there are integers $0 < r_1 < r_2 < r_3 < \dots < r_T$ such that for all $1 \leq v \leq T$ the subsequence $(s_{k-r_v}, \dots, s_{k-r_v+MLK_0-1})$ is an L - M - j -sequence where j is such that

$$\sum_{l=1}^{j-1} t_l < v + c \pmod{T} \leq \sum_{l=1}^j t_l.$$

A k - c -sequence for some c is also called a k -sequence.

We can finally define the steady orbit by assigning a Nash equilibrium to each history $\underline{s} \in (S)^k$. Given such an \underline{s} , apply the following algorithm: start at the end, s_k (the most recent stage) and, going backward, look for an L - M - j -sequence for some j . If no such sequence is found, attach σNE_1 to \underline{s} . Otherwise, assume L - M - j_0 -sequence \underline{s}^1 is found. Continue the search from $s_{MK_0}^1$ (the MK_0 -th component of \underline{s}^1) backward (in \underline{s}), this time looking only for an L - M - j_0 -sequence. If there is none within at most $M(L+1)K_0$ stages before $s_{MK_0}^1$, play σNE_{j_0} if $t_{j_0} > 1$ and $\sigma NE_{j_0 \pmod{m} + 1}$ if $t_{j_0} = 1$.

If, on the other hand, such an L - M - j_0 -sequence \underline{s}^2 was found, continue with it in the same fashion. For $t < t_{j_0}$, if exactly t such sequences are found, play σNE_{j_0} . If t_{j_0} are found, play $\sigma NE_{j_0 \pmod{m} + 1}$.

We contend that any stationary distribution p of this process induces a steady orbit as required. First, it is easy to see that for every $\underline{s} \in (S)^k$, if the process is in state \underline{s} at time t , then, by the choice of K_0 , the probability of being in a k -sequence at time $t+k$ is at least $(1 - \varepsilon_1)$, which means that the stationary distribution probability of these states is at least $(1 - \varepsilon_1)$. Furthermore—again, by choice of K_0 —given any $\underline{s} \in (S)^k$ and every $1 \leq j \leq m$, the probability that σNE_j will be played at least $(t_j L - 1)MK_0$ times during the next $TMLK_0$ stages is at least $(1 - \varepsilon_1)$.

We now wish to show that the stationary distribution probability of all states at which σNE_j is played (according to our algorithm), denoted $p(\sigma NE_j)$, is at least $(t_j/T - \varepsilon_0/m(m-1))$. This would also mean that it is at most $(t_j/T + \varepsilon_0/m)$ and then

$$\begin{aligned} \left\| (1/T) \sum_{j=1}^m t_j NE_j - h(\hat{p}) \right\| &= \left\| (1/T) \sum_{j=1}^m t_j NE_j - \sum_{j=1}^m p(\sigma NE_j) NE_j \right\| \\ &\leq \sum_{j=1}^m |t_j/T - p(\sigma NE_j)| \|NE_j\| < \varepsilon_0. \end{aligned}$$

Let A_j be the set of histories $\underline{s} \in (S)^k$ at which the algorithm above dictates playing σNE_j . Let X_t be the random variable associated with the Markov chain. Then

$$\begin{aligned} p(\sigma NE_j) &= \text{Prob}(X_j \in A_j) \quad (\text{for all } t) \\ &= (1/TMLK_0) \sum_{i=1}^{TMLK_0} \text{Prob}(X_{t+i} \in A_j) \\ &= (1/TMLK_0) \sum_{i=1}^{TMLK_0} \sum_{\underline{s} \in (S)^k} \text{Prob}(X_{t+i} \in A_j | X_t = \underline{s}) p(\underline{s}) \\ &= \sum_{\underline{s} \in (S)^k} p(\underline{s}) (1/TMLK_0) \sum_{i=1}^{TMLK_0} \text{Prob}(X_{t+i} \in A_j | X_t = \underline{s}). \end{aligned}$$

By the previous argument, for each \underline{s} one can find a set of indices $I \subseteq \{1, \dots, TMLK_0\}$ such that $|I| = (t_j L - 1)MK_0$ and that

$$\text{Prob}(X_{t+i} \in A_j, \forall i \in I | X_t = \underline{s}) \geq 1 - \varepsilon_1.$$

Hence, for $i \in I$

$$\text{Prob}(X_{t+i} \in A_j | X_t = \underline{s}) \geq 1 - \varepsilon_1$$

and

$$\begin{aligned} p(\sigma NE_j) &= \sum_{\underline{s} \in (S)^k} p(\underline{s}) (1/TMLK_0) \sum_{i \in I} \text{Prob}(X_{t+i} \in A_j | X_t = \underline{s}) \\ &\geq \sum_{\underline{s} \in (S)^k} p(\underline{s}) (1/TMLK_0) (t_j L - 1) MK_0 (1 - \varepsilon_1) \\ &\geq (t_j/T - 1/LT)(1 - \varepsilon_1) \\ &= t_j/T - 1/LT - \varepsilon_1 t_j/T + \varepsilon_1/LT \\ &> t_j/T - 3\varepsilon_1 = t_j/T - \varepsilon_0/m(m-1), \end{aligned}$$

which completes the proof. ■

Remark 6.4. Notice that, if $k_i = k$ for $i \in N$ and (p, σ) is a steady orbit, then for every player i and every $\underline{s} \in (S)^k$,

$$\tau^i(p, \sigma)(\underline{s})(s_0^{-i}) = \prod_{j \neq i} \sigma^j(\underline{s})(s_0^{-j});$$

that is, player i believes in the “true” σ ’s. The best-response condition then implies that for every $\underline{s} \in (S)^k$, $(\sigma^i(\underline{s}))_{i \in N}$ is a Nash equilibrium of the game G . Hence, for all $k \geq 0$ $SO(k, k, \dots, k) \subseteq Co(NE)$.

Hence, we have proved that

$$\overline{\bigcup_{k \geq 0} SO(k, k, \dots, k)} = Co(NE).$$

PROPOSITION 6.5. Assume that $k_1 \geq k_2 \geq k_3 \geq \dots \geq k_n$. Then $SO(k_1, k_2, \dots, k_n) = SO(k_2, k_2, k_3, \dots, k_n)$. (I.e., the player with the longest recall may restrict himself (herself) to strategies which only depend on histories of length equal to the second-longest recall.)

Proof. Assume (p, σ) is a steady orbit with recall profile (k_1, k_2, \dots, k_n) . Define a steady orbit (p', σ') for the recall profile (k_2, k_2, \dots, k_n) as follows: for $i > 1$, $\sigma'^i = \sigma^i$. For $i = 1$, $\sigma'^1(\underline{s})$ ($\underline{s} \in (S)^{k_2}$) is the p -mixture of $\sigma^1(\underline{s}' \cdot \underline{s})$ (over all $\underline{s}' \in (S)^{k_1 - k_2}$). Since player 1’s best response strategies constitute a convex set, σ'^1 is also a best response to all other players’ strategies given \underline{s} . On the other hand, τ^i for $i \neq 1$ has not changed, so that σ'^i is also a best response strategy. Finally, define $p'(\underline{s})$ as $\sum_{\underline{s}' \in (S)^{k_1 - k_2}} p(\underline{s}' \cdot \underline{s})$. ■

COROLLARY 6.6. For $n = 2$, $\overline{SO} = Co(NE)$.

PROPOSITION 6.7. $SO \subseteq CE$.

Proof. Given a certain recall profile $\{k_i\}_i$ and a steady orbit (p, σ) , we will show that \hat{p} is a correlated equilibrium. Hence, perforce, $h(\hat{p}) \in CE$. For some $s \in S$ with $\hat{p}(s) > 0$ and $i \in N$, we have to show that s^i is a best response move for player i while the other players are playing according to

$$\frac{\hat{p}(\cdot, s^i)}{\sum_{t^i \in S^{-i}} \hat{p}(t^i, s^i)} \in \Delta(S^{-i}).$$

Note that

$$\begin{aligned}
 \hat{p}(t^{-i}, s^i) &= \sum_{\underline{s} \in (S)^{k-1}} p(\underline{s} \cdot (t^{-i}, s^i)) \\
 &= \sum_{\underline{s} \in (S)^{k-1}} \sum_{\underline{s}' \in (S)^k} p(\underline{s}') a(\sigma)_{(\underline{s}', \underline{s} \cdot (t^{-i}, s^i))} \\
 &= \sum_{\underline{s} \in (S)^{k-1}} \sum_{\underline{r}_1 \in (S)^{k_i}} \sum_{\underline{r}_2 \in (S)^{k-k_i}} p(\underline{r}_2 \cdot \underline{r}_1) a(\sigma)_{(\underline{r}_2, \underline{r}_1 \cdot \underline{s} \cdot (t^{-i}, s^i))} \\
 &= \sum_{\underline{s} \in (S)^{k-1}} \sum_{\underline{r}_1 \in (S)^{k_i}} \bar{p}(\underline{r}_1) \sum_{\underline{r}_2 \in (S)^{k-k_i}} \bar{p}(\underline{r}_2) a(\sigma)_{(\underline{r}_2, \underline{r}_1 \cdot \underline{s} \cdot (t^{-i}, s^i))}
 \end{aligned}$$

where

$$\bar{p}(\underline{r}_1) = \sum_{\underline{r}_2 \in (S)^{k-k_i}} p(\underline{r}_2 \cdot \underline{r}_1)$$

and

$$\bar{p}(\underline{r}_2) = p(\underline{r}_2 \cdot \underline{r}_1) / \bar{p}(\underline{r}_1) \text{ or zero if } \bar{p}(\underline{r}_1) = 0.$$

By definition of τ^i ,

$$\sum_{\underline{s} \in (S)^{k-1}} \sum_{\underline{r}_2 \in (S)^{k-k_i}} \bar{p}(\underline{r}_2) a(\sigma)_{(\underline{r}_2, \underline{r}_1 \cdot \underline{s} \cdot (t^{-i}, s^i))} = \sigma^i(\underline{r}_1)(s^i) \cdot \tau^i(p, \sigma)(\underline{r}_1)(t^{-i})$$

whence

$$\hat{p}(t^{-i}, s^i) = \sum_{\underline{r}_1 \in (S)^{k_i}} \bar{p}(\underline{r}_1) \sigma^i(\underline{r}_1)(s^i) \tau^i(p, \sigma)(\underline{r}_1)(t^{-i})$$

and

$$\sum_{t^{-i} \in S^{-i}} \hat{p}(t^{-i}, s^i) = \sum_{\underline{r}_1 \in (S)^{k_i}} \bar{p}(\underline{r}_1) \sigma^i(\underline{r}_1)(s^i).$$

Combining these equalities, the conditional distribution of player i on the other players' moves (the ratio of the last two expressions) is a convex combination of $\{\tau^i(p, \sigma)(\underline{r}_1)\}_{\underline{r}_1 \in (S)^{k_i}}$. However, only $\underline{r}_1 \in (S)^{k_i}$ for which $\sigma^i(\underline{r}_1)(s^i) > 0$ have a positive coefficient. Note that for such \underline{r}_1 the move s^i has to be a best response to $\tau^i(p, \sigma)(\underline{r}_1)$. Hence s^i is also a best response to the convex combination of these distributions. ■

REMARK 6.8. *SO is not necessarily convex.*

Proof. Consider the three-person game

		Player II				Player II	
		F	S			F	S
Player I	F	0.1, 0.1, 0.1	10, 0, 10	Player I	F	10, 10, 0	0, 0, 0
	S	0, 10, 10	0, 0, 0		S	0, 0, 0	1, 1, 1
		F				S	

(Player III chooses the matrix.) For a recall profile $(9, 9, 0)$ (that is, $k_1 = k_2 = 9, k_3 = 0$), the following is a (pure) steady orbit: Players I and II play S if they have observed (F, F, S) in the last nine periods, and F otherwise. Player III always plays S .

Thus, the steady orbit play consists of nine (F, F, S) and one (S, S, S) repeated cyclically. For Players I and II the strategy is obviously a best response one. Player III observes that the other players play (F, F) with 90 percent frequency and (S, S) otherwise. Hence his/her strategy is also a best response one. The average payoff vector is $(9.1, 9.1, 0.1)$.

Similarly, $(9.1, 0.1, 9.1)$ and $(0.1, 9.1, 9.1)$ are also in SO (for different recall profiles). If SO were convex, we would have to conclude that $(6.1, 6.1, 6.1)$ is also in SO . Let us prove that this is impossible. Suppose, then, that (k_1, k_2, k_3) is a recall profile for which there is a steady orbit (p, σ) such that $h(\hat{p}) = (6.1, 6.1, 6.1)$ and assume, without loss of generality, that $k_1 \geq k_2 \geq k_3$. Consider $H(s) = h^1(s) + h^2(s) + h^3(s)$ and extend H linearly to $\Delta(S)$. Note that $H(\hat{p}) = 18.3$ and that the only three points $s \in S$ for which $H(s) \geq 18.3$ are (F, F, S) , (F, S, F) , and (S, F, F) . For those $H(s) = 20$. The maximal value of $H(s)$ for other points s is 3, hence $\hat{p}((S, F, F)) + \hat{p}((F, S, F)) + \hat{p}((F, F, S)) \geq 0.9$. If ε satisfies $\hat{p}((S, F, F)) < \varepsilon$, then $h^1(\hat{p}) \geq 9 - 10\varepsilon$ whence $\hat{p}((S, F, F)) \geq 0.29$. Hence there must be $\underline{s} \in S^k$ such that the probability of (S, F, F) in the next move is at least 0.29. Suppose that at this node (\underline{s}) Player II plays S with probability p and Player III—with probability q . Since $k_1 = k$, player I knows these probabilities. For him/her to play S with positive probability the following inequality should hold:

$$(1 - p)(1 - q) \geq 0.1pq + 10(1 - p)q + 10(1 - q)p.$$

Some algebra shows that this is impossible if $pq \geq 0.29$.

(The above inequality means that

$$20.9 pq + 1 \geq 11(p + q).$$

Since $pq \geq 0.29$ this means that $(p + q) \leq 7.061/11 \leq 0.65$, for such p , $q \max pq \leq 0.11$.) ■

Remark 6.9. The above example can also be used to show that \overline{SO} is not necessarily convex. Hence \overline{SO} is *strictly* included in CE .

7. CONCLUDING REMARKS

7.1. It has been argued that steady orbits, with the unavoidable interpretation of repeated games, are not robust with respect to partitions of the time periods: the infinitely repeated game may be thought of as an infinite repetition of k -repetitions of the one-shot game for some $k > 1$. In this case the set of actions will be larger, and so will the set of equilibria payoffs; in fact, one may get the Folk Theorem.

There is, however, a crucial difference between $k = 1$ and $k > 1$: for $k = 1$ it is reasonable to assume that each player's action is observed by the others. The strategy in a repeated game cannot be observed in the same way.

Admittedly, by the same logic one concludes that steady orbits make more sense for a one-stage simultaneous move game than for general extensive form games. At least for this subclass of games, which may successfully model a wide range of interaction situations, we find steady orbits to be a viable solution concept.

7.2. The study presented above may be viewed as an attempt to formulate the "repeated game" interpretation of Nash equilibrium in the one shot game: a possible motivation for this concept (which is quite often used) is that if the game is repeated, and should a certain play of it be constantly chosen, this play must be a Nash equilibrium. Indeed, this intuition is reinforced by our results if all the players have zero memory length. However, the bounded rationality arguments only imply bounded (and not necessarily equal) memory lengths. Thus, we have found that a larger set of payoffs—namely, SO —may be justified on the same grounds.

7.3. In order to distinguish between real numbers as people seem to perceive them and real numbers as encoding of extremely complicated strategies we used m - T -Turing machines, having finitely many real-valued registers, and restricted to use them only a bounded number of times in each computation. This sufficed for the bounded recall result to hold (Observation 5.1), and then the behavioral assumption of nonstrategic

players specified what the players will actually do with their memory registers.

However, we are not quite happy with the computational model presented here: while it restricts the complexity of each single computation, the sequence of moves a player chooses in the repeated game may still be quite complicated. For instance, a player may decide to encode his/her opponent's moves during N periods and then play according to some function of them (say, "tit-for-tat") for the next N periods. It is easy to see that this strategy does not require any recall at all, and that only one register, which is used once in each computation, suffices to implement it.

A natural way to solve this problem is to allow only finite memory instead of real-valued registers. Thus, the memory may contain approximations of real numbers, but cannot be infinitely complex. We rejected this solution because it does not draw the intended distinction: the precision of the approximation will also determine the complexity of the strategy.

Another solution may be to simply define a computational model that can do exactly what we want it to do, namely, update the appropriate statistics. Of course, this is a very restrictive model.

It is therefore left as an open problem to find a general computational model that distinguishes between the way people and machines handle numbers.

7.4. Finally, we note that one may obtain similar results with various versions of the assumptions: one may use a game with "stage 0" and explicitly assume (rather than deduce) bounded recall strategies; one may use Turing machines with finite memory, and so forth. However, we find the version presented here the most satisfactory from a conceptual viewpoint.

REFERENCES

- ABREU, D., AND RUBINSTEIN, A. (1988), "The Structure of Nash Equilibrium in Repeated Games with Finite Automata," *Econometrica* **56**, 6, 1259-1281.
- AUMANN, R. J. (1974), "Subjectivity and Correlation in Randomized Strategies," *J. Math. Econ.* **1**, 67-95.
- AUMANN, R. J. (1981), "Survey of Repeated Games," in *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern*, Mannheim/Weir/Zurich: Bibliographisches Institut.
- AUMANN, R. J., AND SORIN, S. (1989), "Cooperation and Bounded Recall," *Games Econ. Behav.* **1**, 5-39.
- BEN-PORATH, E. (1993), "Repeated Games with Finite Automate," *J. Econ. Theory* **59**, 17-32.
- BINMORE, K. (1986), "Modeling Rational Players (I, II)," *J. Econ. and Phil.* to appear.

- GILBOA, I. (1988), "The Complexity of Computing Best-Response Automata in Repeated Games," *J. Econ. Theory* **45**, 342–352.
- GILBOA, I., AND SAMET, D. (1989), "Bounded Versus Unbounded Rationality: The Tyranny of the Weak," *Games Econ. Behav.* **1**, 213–221.
- GILBOA, I. (1990), "Philosophical Applications of Kolmogorov's Complexity Measure," mimeo.
- HOPCROFT, J. E., ULLMAN, J. D. (1979), *Introduction to Automata Theory, Languages and Computation*, Reading, MA.
- KALAI, E., AND STANFORD, W. (1985), "Equally Sophisticated Players: On the Complexity, Memory and Automation of Repeated Game Strategies," manuscript.
- KALAI, E., STANFORD, W. (1988), "Finite Rationality and Interpersonal Complexity in Repeated Games," *Econometrica* **56**, 2, 397–410.
- LEHRER, E. (1988a), "Repeated Games with Stationary Bounded Recall Strategies," *J. Econ. Theory* **46**, 130–144.
- LEHRER, E. (1988b), " n Players with Bounded Recall in Infinitely Repeated Games," manuscript.
- MEGIDDO, N., AND WIGDERSON, A. (1986), "On Play by Means of Computing Machines," manuscript, Northwestern University.
- NASH, J. (1951), "Non-cooperative Games," *Annals Math.* **54**, 286–295.
- NEYMAN, A. (1985), "Bounded Complexity Justifies Cooperation in the Finitely Repeated Prisoner's Dilemma," *Econ. Lett.* **19**, 227–229.
- PREPARATA, F. P., AND SHAMOS, M. I. (1985), *Computational Geometry: An Introduction*, New York: Springer-Verlag.
- RUBINSTEIN, A. (1986), "Finite Automata Play the Prisoner's Dilemma," *J. Econ. Theory* **39**, 83–96.
- SCHWARTZ, G. (1974), "Randomizing When Time is not Well-Ordered," *Israel J. Math.* **19**, 241–245.
- SIMON, H. A. (1972), "Theories of Bounded Rationality," in *Decision and Organization* (C. B. McGuire and R. Radner, Eds.). Amsterdam: North-Holland.
- SIMON, H. A. (1978), "On How to Decide What to Do," *Bell J. Econ.* **9**, 494–507.
- STANFORD, W. (1987), "Some Simple Equilibria in Repeated Games," manuscript.