AUTOMATA AND INNER STATES FOR REPEATED GAMES

The repeated Prisoner's Dilemma is often used as a game-theoretic paradigm to analyse the evolution of reciprocity, of mutual aid, and of trade. Conceivably, it can also help in understanding the emergence of inner motivational states, like contribution or outrage, which are essential for fuelling the ethics of communal life. A highly internalised sense of fairness and the readiness for moralistic aggression, the ability to be provoked and the feeling of guilt are important and apparently ubiquitous aspects of human socialisation. Like Ridley (1996) and Frank (1988) we believe that emotional commitment is a major factor in the economics of every-day life.

Let us consider a two-player game where both players have the same two strategies and the same payoff matrix. We denote the first strategy by C (for 'cooperate') and the second by D (for 'defect') and write the payoff matrix as

\[
\begin{pmatrix}
R & S \\
T & P
\end{pmatrix}
\]

Such games include the Prisoner's Dilemma, where \( T > R > P > S \), and the Chicken game, where \( T > R > S > P \). (In the Prisoner's Dilemma case, \( R \) stands for the reward for mutual cooperation, \( P \) is the penalty for mutual defection, \( T \) is the temptation payoff for unilaterally defecting and \( S \) the sucker payoff for being exploited. The strategy \( D \) dominates \( C \), so that rational players will defect and thus earn only \( P \) instead of \( R \). One also assumes \( 2R > T + S \).)

Let us assume now that the game is repeated with a constant probability \( w \). The number of rounds is a random variable with expected value \((1-w)^{-1}\). The total payoff is given by \( \sum A_n w^n \), with \( A_n \) as payoff in the \( n \)-th round and \( w^n \) the probability for the occurrence of an \( n \)-th round. In the limiting case \( w = 1 \) (the infinitely iterated game) one uses as payoff the limit in the mean, i.e. \((A_1 + \ldots + A_n)/n\) (provided it exists). For the PD game and \( w \) sufficiently large, there exists now no strategy which is best against all comers (see Axelrod, 1984). For \( w > (T - R)/(T - P) \), for instance, the best reply against AlwaysC is to always defect, whereas against Grim (the strategy that cooperates up to the first time it has been exploited, and then always defects) it is best to always cooperate.

The so-called folk theorem for repeated games implies that there exist infinitely many Nash-equilibria (see e.g. Fudenberg and Tirole 1992 or Binmore 1994) – in particular, every feasible pair of payoff-values larger than the maximin value can be attained by Nash-equilibria. On the other hand, it is easy to see that there exist no strict Nash equilibria, and in fact no evolutionarily stable strategies in the strict sense (see Binmore and Samuelson 1982; with a different notion of evolutionary stability, see Lorberbaum 1994). A strategy \( E \) is said to be evolutionarily stable.
stable if in addition to the Nash-equilibrium condition $A(E', E) \leq A(E, E)$ one has $A(E', E') < A(E, E')$ for all strategies $E \neq E'$ (here $A(E', E)$ is the payoff for a player using strategy $E'$ if the co-player uses $E$). This implies that a population of $E$-players cannot be invaded by a minority of $E'$-players under the effect of natural selection. (See Maynard Smith, 1982, and Hofbauer and Sigmund, 1988).

It is easy to see, for example, that the strategy Tit For Tat (or TFT) is not evolutionarily stable (although this has been claimed occasionally). TFT is the strategy which plays C in the first round and from then on repeats whatever the co-player did in the previous round. This strategy did very well in a series of computer tournaments (see Axelrod 1984). But a player using Always C does as well, both against TFT and against its own kind, as TFT. Intuitively speaking, in a population of TFT-players, invading Always C-players can spread by neutral drift, ultimately allowing defectors to cash in.

Another weakness of TFT is that it suffers from occasional errors against its like, as has been pointed out by Selten and Hammerstein (1982) already. If, by some mistake, a TFT-player plays D against another TFT-player, this leads to a long vendetta of alternating unilateral defections which can only be stopped by a further mistake: such a second mistake, however, can lead just as well to a regime of simultaneous defections. This lowers the average payoff in a TFT-population considerably and allows more generous strategies to invade.

In Nowak and Sigmund (1993, 1995) evolutionary chronicles have been studied by computer simulations of large populations of players using strategies defined by the propensities $p_R, p_S, p_T$ and $p_P$ to play C after having experienced the payoff $R, S, T$ resp. $P$ in the previous round (and, for $w < 1$, the probability to play C in the first round). We note in passing that Boerlijst et al. (1997) have shown that in case $w = 1$, there exist for every value $\pi$ between $P$ and $R$ uncountably many strategies of this type with the property that every co-player obtains $\pi$ as payoff, no matter which strategy he is using. In particular, for every pair of values $\pi$ and $\psi$ between $P$ and $R$, this yields Nash-equilibria composed of pairs of such ‘equaliser’ strategies, and provides a variant of the folk theorem.

The evolutionary simulations lead frequently to cooperative regimes dominated by the Pavlov strategy, the strategy with $p_R = p_P = 1$ and $p_S = p_T = 0$. This strategy, which is based on the win-stay, lose-shift principle of repeating the former move if and only if it leads to a high payoff, i.e. to $R$ or to $T$, has the property of being error-correcting. If, in a game between two Pavlov-players, one erroneously plays D, then in the next round both play D and after that resume mutual cooperation.

Sugden (1988) has proposed another, more sophisticated strategy called Con- trite TFT (cTFT). It is based on the notion of standing, which is associated to each player and can be $g$ (good) or $b$ (bad). In each round, the player acts (i.e. opts for C or D) and obtains a new standing which depends on his action and on the previous standing of both players. The rules for updating the standing are the following: if the other player has been in good standing, or if both have been in bad standing, one receives a good standing if one has cooperated, and a bad standing otherwise. If one has been in good standing and the other player in bad standing, one receives
a good standing no matter what one is doing.

Thus if one cooperates in a given round, one will always obtain a good standing: but if one defects, one will be in good standing only if one has been in good standing and the opponent in bad standing in the previous round.

In a given round, a player can be in three possible states: $Cg$, $Dg$ and $Db$: the first means that he has cooperated (which automatically entails good standing), the second that he has defected with good reason (i.e. while ‘provoked’), the third that he has wantonly defected and feels the pangs of conscience. The state of the game in a given round is made up of the states of the first and the second player. There are nine such combinations: $(Cg, Cg)$, $(Cg, Dg)$, $(Cg, Db)$, $(Dg, Cg)$, $(Dg, Db)$, $(Db, Cg)$, $(Db, Dg)$, $(Db, Db)$ and $(Dg, Dg)$. It is easy to check that this last state can never be reached: we therefore omit it, and number the remaining eight states in this order.

$cTFT$ is the strategy which cooperates except if it is in good standing and the other player is not. This means that the player defects when provoked, but not otherwise. If he defects by mistake, he knows that he lost his good standing, and meekly accepts punishment, i.e. keeps cooperating even if the other player uses $D$ on him.

In other words, the strategy $cTFT$ begins with a cooperative move, and cooperates except if provoked (or by mistake). If two $cTFT$-players engage in a repeated Prisoner’s Dilemma, and the first player defecets by mistake, then he loses his good standing. In the next round, he will cooperate, whereas the other $cTFT$-player will defect without losing his good standing. From then on both players will be in good standing and resume their mutual cooperation in the following round.

In Boerlijst et al (1997) we have studied evolutionary chronicles for strategies defined by their propensities $(q_1, \ldots, q_8)$ to play $C$ when in one of the eight states previously enumerated. This large class contains the $(p_R, p_S, p_T, p_P)$-strategies in the form $(p_R, p_S, p_S, p_T, p_P, p_T, p_P, p_P)$. The strategy $TFT$ is for instance given by $(1, 0, 0, 1, 0, 1, 0, 0)$, the strategy $Pavlov$ is $(1, 0, 0, 0, 1, 0, 1, 1)$ and the rule for $cTFT$ is $(1, 1, 0, 1, 0, 1, 1, 1)$.

These chronicles show that $cTFT$ works very well indeed. We may distinguish two situations. If the temptation $T$ is large (for instance, $T = 5.5$, $R = 3$, $P = 1$, $S = 0$) then the simulations end up in about 80 percent of all runs with a population dominated by $cTFT$, and in most other runs with a population dominated by $Remorse$. $Remorse$ is the strategy $(1, 0, 0, 0, 1, 0, 1, 1)$ which cooperates only if in bad standing, or if both players had cooperated in the previous round.

If on the other hand the temptation to defect is small (for instance $T = 3.5$, and the other values as above) then 70 percent of all runs end up with $cTFT$ or some similar strategy, 20 percent with $Pavlov$ and the rest with a strategy called $Weakling$, which plays $C$ if and only if it is in bad standing, and which obtains against itself a sub-Pareto optimal payoff (namely $(R + P)/2$ for $w = 1$) by simultaneously cooperating and defecting in every other round. It is surprising that this suboptimal trapping is only observed in situations with low temptation to defect.

In contrast to $Pavlov$, which does very poorly in a population of $AlwaysD$ players (it plays $C$ in every second round), $cTFT$ and $Remorse$ are as good at invading an $alwaysD$-population as are the strongly retaliatory strategies $TFT$ or $Grim$. 
This is one of the reasons for the success of \textit{cTFT}; it does not need a strategy which catalyses cooperation, as \textit{Pavlov} does.

In order to better understand why the outcomes described in the previous simulations are stable, we can use a minor variant of the concept of limit-ESS introduced by Selten (see Selten, 1984, and Leimar, 1997). Let us first note that the description of a repeated game by its extensive form suffers from the fact that there is no uniform upper bound for the length of the branches of the game-tree. If we consider only finite state automata, this problem is overcome, since we have only to investigate a finite number of nodes. This yields a ‘closure’ of the representation of the repeated game.

For simplicity let us consider first only strategies depending on the moves of the previous round, like \textit{TFT} or \textit{Pavlov}. In the repeated game, there occur only four pairs of states, namely (C, C), (C, D), (D, C) and (D, D). The history of every repeated game corresponds to a path connecting these four vertices. Let us consider the fate of a player invading a population of \textit{TFT} players, for instance (see Fig. 1).

\begin{figure}
\centering
\includegraphics[width=0.8\textwidth]{figure1.png}
\caption{Figure 1}
\end{figure}

\begin{figure}
\centering
\includegraphics[width=0.8\textwidth]{figure2.png}
\caption{Figure 2}
\end{figure}

In every state, the move of the co-player, who belongs to the resident population, is specified by the \textit{TFT} rule. The invader has two alternatives. He can use the move prescribed by the \textit{TFT}-rule, which we describe in the graph by a full arrow, or he can use the other option, described by the broken arrow. One sees immediately that it is better, in the state (C, D), not to use \textit{TFT}, since this would yield as outcome for the next two rounds only $T + wS$; it would be better to cooperate and thus to reach (C, C) since this yields (for the next two rounds) $R + wR$. This latter option is better provided $w$ is sufficiently large. We note that in a game between two \textit{TFT} players the node (C, D) can only be reached by mistake. According to Selten’s theory of perfect equilibria, such states cannot be neglected.
Let us consider, in contrast, a player invading a Pavlov-population (see Fig. 2). In this case, one sees that as long as $T + wP < R + wR$, it is best to do, in every node, as the residents do. If the converse inequality holds, it is better to deviate in the states $(C, C)$ and $(D, D)$. We say that for $T + wP < R + wR$, Pavlov is a limit ESS (see Wu and Axelrod, 1995).

More generally, we say that a strategy $E$ implemented by a finite-state automaton is limit-ESS if for every pair of states which can be reached in the normal course of a game between two $E$-players or after mis-implementations, it is better to follow the $E$-rule than the alternative (i.e. the full arrow yields always the path with the highest payoff). In Fig. 3 we check that Remorse is a limit-ESS if $T + wP > R + wR$ (just the opposite as with Pavlov) whereas Fig. 4 shows that $cTFT$ is always a limit ESS. Olof Leimar (1997) has developed an extensive theory of limit-ESS for state space strategies, and has shown that for automata with three or four states, there exist thousands of such strategies which (like Pavlov, Remorse or $cTFT$) are Pareto-optimal for large $w$.

Figure 3

On the other hand, while $cTFT$ is immune to errors in the implementation of a move, it is not immune to errors in the perception of a move. If, in a match between two $cTFT$ players, one player mistakenly believes that the other is in bad standing, this leads to a sequence of mutual backbiting, just as with $TFT$. In contrast to this, if
in a game between two Pavlov-players one player mis-interprets a move (by his co-player or by himself), then the situation is quickly redressed and mutual cooperation resumed after two rounds.

Errors in perception – rather than implementation – have been studied in Nowak et al. (1995), among others. In order to see that they are really different from errors in implementation, let us return to the class of \((pR,pS,pT,pP)\)-strategies and assume that the error probability is \(\epsilon\). If this is an error in implementing a move (i.e. in choosing C or D) then \(pR\) turns into \((1-\epsilon)pR+\epsilon(1-pR)\) etc. so that the correction term is \(\epsilon(1-2pR,1-2pS,1-2pT,1-2pP)\). If the error affects the perception of the co-player’s move (i.e. if it confuses an \(R\) with an \(S\) or a \(T\) with a \(P\)) then the correction term is \(\epsilon(pS-pR,pR-pS,pP-pT,pT-pP)\). If the error affects the perception of the own move (i.e. if it confuses an \(R\) with a \(T\) or an \(S\) with a \(P\)) then the correction term is \(\epsilon(pT-pR,pP-pS,pR-pT,pS-pP)\). For \(w=1\) and the limit \(\epsilon \to 0\), errors in implementation yield as payoff \((2P+2S+T)/5\) for a \((1,0,0,0)\)-player against a \((0,0,1,0)\)-player, whereas errors in perceiving the opponent’s move yield as payoff \((S+T)/2\).

As soon as we allow mistakes in perception, we lose the concept of a public state (like the standing). In the elementary set-up we are considering, there is no referee to tell the players who is in which state. Every player can only monitor his own state. This state need not be restricted to his own previous move, or to his own standing.

For instance, we can realise the cTFT strategy by an automaton with three states 1, 2, 3. State 2 is \((b,g)\) (my standing is bad, the co-players good); state 3 is \((g,b)\); and state 1 is \((g,g)\) or \((b,b)\). The cTFT rule then is to play C when in state 1 or 2, and D when in state 3. Outcome \(R\) (mutual cooperation) leads always to state 1, outcome \(S\) leads from states 1 or 3 to 3 and from state 2 to 1, outcome \(T\) leads from 1 and 2 to 2 and from 3 to 1, and outcome \(P\) leaves all states unchanged. If two cTFT-players make no mistakes in interpretation, then they will either both be in state 1, or one player in 2 and the co-player in 3 (see Leimar, 1997). But misperceptions can change this symmetry, and lead to vendettas.

The trembling hand doctrine of Selten assumes that the player is completely lucid. He can make a wrong move, but he is immediately aware of his gaffe. Such a player is still too rational for the kind of evolutionary games we have in mind. A player is also liable to errors in interpreting the outcome \(R, S, T\) or \(P\) of the previous round, and in updating his state as a function of this outcome.

This suggests considering strategies implemented by automata of the following kind. Each consists of a finite set \(\Omega\) of inner states and a (possibly stochastic) action rule with the probability \(p_\omega\) for the player to choose the move C when in state \(\omega \in \Omega\). In addition, there exists a (possibly stochastic) transition rule \(\rho\) which specifies the state in the next round, depending on the current state and the outcome \(R, S, T\) or \(P\) of the current round. If we allow mistakes in implementing a move, \(p_\omega\) will not attain the extremal (deterministic) values 1 oder 0, and each outcome \(R, S, T\) or \(P\) can occur with positive probability. If we assume mistakes in the interpretation, this implies that the transition rule \(\rho: \Omega \to \Omega\) is also stochastic. To specify the strategy completely, we also need the initial state (i.e. a probability distribution \(q\) on \(\Omega\)).
The actual repeated game between two automata \((\Omega, p, \rho, q)\) and \((\Omega', p', \rho', q')\) is described by a path visiting the vertices of the state space \((\Omega, \Omega')\) of the game. The transitions within this state space are described by a Markov chain. When in state \((\omega, \omega')\), the outcome of the next round, i.e. \(R, S\) etc. is given by the probabilities \(p_{\omega}p'_{\omega'}, p_{\omega}(1 - p'_{\omega'})\) etc. and this outcome leads with the probabilities defined by \(\rho\) and \(\rho'\) to the follow-up states. Let us denote the transition probability from the state \((\omega, \omega')\) to the state \((\alpha, \alpha')\) by \(P(\omega, \omega'; \alpha, \alpha')\).

If one starts in the state \((\omega, \omega')\), the first player obtains a total payoff whose expected value is \(E(\omega, \omega')\). It is easy to compute these values by using the fact that in the next round (which occurs with probability \(w\)) the whole process starts as if from scratch (it is a renewal process). Hence

\[
E(\omega, \omega') = R p_{\omega} p'_{\omega'} + S p_{\omega} (1 - p'_{\omega'}) + T (1 - p_{\omega}) p'_{\omega'} + P (1 - p_{\omega}) (1 - p'_{\omega'}) + w \sum_{\alpha, \alpha'} E(\alpha, \alpha') P(\omega, \omega'; \alpha, \alpha').
\]

This yields a finite system of linear equations for the \(E(\omega, \omega')\). We only have to multiply these values with the probabilities \(q_{\omega} q'_{\omega'}\) for the initial states \((\omega, \omega')\), and obtain by summation the expected total payoff for the automaton \((\Omega, p, \rho, q)\) against \((\Omega', p', \rho', q')\).

We can formulate the notion of limit-ESS in this context too. For an automaton to be a limit-ESS, we must require that in every state which can be reached by the automaton playing against itself (either in the normal run of the game or after a mistake), both the prescribed move and the prescribed transition are the unique optimal moves. This implies immediately that such rules have to be deterministic – for otherwise, two alternatives would fare equally well. It seems improbable that there exist Pareto-optimal strategies which are in this sense limit-ESS. It does not suffice that the strategy is immune against errors in perception and implementation (as for instance Pavlov if \(T + wP < R + wR\)). If one Pavlov-player is for instance in state \((C, D)\) (and believes himself to have been exploited) whereas the other Pavlov-player is in state \((C, C)\), then both will be in the next round in state \((D, C)\), in the following round in state \((D, D)\), and only from then on in \((C, C)\). It would obviously have been better if the first player had chosen \(C\) right away. But this would have been the wrong move if the co-player had shared the same view of the situation with the first player.

Evolutionary simulations in the world of such automata should be quite interesting. Even in the very restricted case of two-state automata, this could well produce new perspectives, since most simulations so far did not take errors of perception into account. But the real challenge lies in automata with more states. If strategies like cTFT emerge, this would be a strong hint that inner states like guilt or wrath serve an economic purpose, and can be viewed as instruments to make our societies tick. One way to design simulations able to introduce new structures among automata – new states, for instance – has been shown by Lindgren (1993), and mimicks gene duplications. One mutation could duplicate a state \(\omega\): the resulting two states would
act in the same way, so that no new behaviour emerges. Such a mutation is ‘silent’ and selectively neutral. In a second step, other mutations could modify the action rules and the transition rules in the two states in different ways. These mutations would now be able to differentiate the two states, and hence to build new, more complex automata.

Lindgren (1991) used this principle to investigate strategies dependent on the moves of the last two, three, four rounds etc. cTFT and Remorse are strategies of a different nature. They only depend on the state of the previous round, but this state, now, does not depend only on the actions C or D of the two players, but on the standing – good or bad – after a defection. The rules for determining this standing seem quite natural: we can identify with a player who feels bad after having committed a defection erroneously, or who feels provoked by the unilateral defection of the co-player after a string of mutual cooperation. The rules embody a certain notion of ‘fairness’ which seems to be rather common. If it should indeed turn out that this notion is a human universal, we would have to explain how it emerged.

It seems highly plausible that there exists a wide variety of workable ‘taggings’ or states which yield interesting ESS’s. The question is whether an evolution based on mutation and selection would tend to lead to one form of states rather than another. This could ultimately shed light on why humans developed a sense of fairness, feelings of guilt, and highly effective social norms. So far, however, this is science fiction.

REFERENCES


Karl Sigmund
Institut für Mathematik, Universität Wien
Strudlhofgasse 4, A-1090 Vienna, Austria

Maarten Boerlijst
Department of Zoology, University of Oxford
South Parks Road, Oxford OX1 3PS, U.K.

Martin A. Nowak
Department of Zoology, University of Oxford
South Parks Road, Oxford OX1 3PS, U.K.