

Cycling Co-Evolution Resulting from Genetic Adaptation in Two-Person Zero-Sum Games

David Salamon and Peter Salamon

Department of Mathematical and Computer Sciences

San Diego State University

San Diego, CA 92182-7720

Phone:619-594-7204

salamon@sdsu.edu

Running Title: Cycling Co-Evolution in Two-Person Zero-Sum Games

Abstract

We consider two populations co-evolving with fitness defined by the payoff in a two-person zero-sum game. We show that such situations lead to spontaneous and sustained oscillations iff the optimal strategy of the game is mixed.

KEYWORDS: co-evolution, zero-sum games, evolutionary games

Introduction

Consider two populations co-evolving in the following manner: each member of population $\mathcal{A}=\{A\}$ plays a certain 2-person zero-sum game repeatedly with random members of population $\mathcal{B}=\{B\}$. Each player uses a fixed mixed strategy in all the games in which he participates. After a certain number of games by each player, a new generation is created according to an evolutionary strategy whose fitness function is the accumulated winnings or losses of the player. The evolutionary process selects for the individuals with better strategies. This short paper concerns a general observation regarding the course of evolution in such a scenario: When the optimal strategies in the associated game are mixed, the populations spontaneously oscillate around the game theoretically optimal strategies. This observation, once made, becomes obvious. Nonetheless, it does not seem to have been noted in the literature, and has interesting and far-reaching implications.

A number of authors (Maynard-Smith, 1982; Axelrod, 1987; Boyd and Lorberbaum, 1987; Rovatsos and Lind 1999; Darwen and Yao, 1995; Yao and Darwen, 1994) have reported on the use of evolutionary strategies for exploring cooperative behavior in iterated non-zero-sum games, most notably the iterated prisoner's dilemma. The aim of these numerical experiments is to explore the extent to which the time evolution of the resulting strategies can model the formation and stability of human cooperation. Co-evolving strategies have also been a popular topic (Maynard-Smith 1982; Rosin and Belew, 1996; Haynes and Sen, 1996; Akiyama and Kaneko, 1995) under the rubric of *evolutionary games*. The fact that co-evolution in zero-sum games with pure strategies converges to the optimal solution has been noted (Tizard, 2000; Holland, 1975; van Nimwegen et al., 1997) and even its method of convergence by way of punctuated equilibria has been charted (van Nimwegen et al., 1997). Here, as in most of the literature on game theory, the focus has been on convergence to the game-theoretically optimal strategies known as evolutionary stable strategies (ESS).

The fact that co-evolution in zero-sum games with mixed optimal strategies does not converge to the optimum but rather *undergoes a spontaneous transition to cyclic behavior* has not been noted. This may be due to the fact that the attention in this area has focused primarily on non-zero-sum games or zero-sum games with pure strategies. Further, the focus has been on the behavior of "rational" agents that try their best to reason ahead from the information given. Our goal is different. We explore the use of genetic algorithms among two populations of players playing an iterated 2-person zero-sum game as a model for co-evolving strategies by two populations with limited

adaptation. The limitation on the adaptation is that any genetic operators (Goldberg, 1989; Haupt and Haupt, 1998) (selection, mutation, crossover) act on all members of the populations equally based only on their current fitness in relation to the current fitness of others in the same population. Moreover, such fitness is defined by a strictly monotonic function of the accumulated winnings or losses in the game. We show that this family of algorithms leads to cycles iff the optimal strategy in the game is a mixed strategy. The implications for economic interactions, biological evolution, and artificial intelligence are briefly explored.

The basic reason for the cycling behavior follows from the "hedge all your bets" nature of the optimal strategy in a 2-person zero-sum game. This means that when player A uses the game-theoretically optimal strategy, he has hedged himself against the worst player B can do. In this case the complementary slackness portion of the Kuhn-Tucker conditions assure us that the outcome of the game will be *independent* of the mixed strategy player B adopts, provided he only mixes strategies that are part of the blend in his game-theoretically optimal strategy. In an adaptive setting, notably a setting with random fluctuations, this means that the population of B players can drift a long way from their optimal strategy without feeling any correlation between fitness and strategy among the B players. Not so for the population of A players. If B has deviated from the hedged all bets position, the A players' fitness distribution will correlate with certain modifications in the population that take advantage of the B players' suboptimality. This causes the B players to evolve back towards optimality while the A players continue trying to exploit B's suboptimal position. Because the A players' exploitation is itself a deviation from the

hedged strategy, this causes the B players to evolve beyond their optimal strategy trying to cash in on the A players' deviation from equilibrium and this leads to growing oscillations of the strategies around the optimum.

It is certainly possible to use evolutionary algorithms to find optimal mixed strategies (Koza, 1992; Holland, 1975; Fogel, 1993). These algorithms typically adapt one population at a time comparing how each strategy performs against all (or a large *fixed* sample) of the other player's possible strategies and choosing the best alternative. The cyclic nature of *co*-evolution for mixed solution games has not been noted in these efforts either. Our co-evolving algorithm does provide a way to find the game-theoretic optimum. In our version of the algorithm, we can simply average the strategies over many cycles. Such averages converge to the game theoretic optima. Any techniques based on evolutionary algorithms will however fail to give linear programming serious competition in games where the payoff matrix is known since the computational efforts required by evolutionary techniques are much greater.

An Example

Consider the classic game of Pennies described for example by Vajda in 1960, where two players each place a coin either heads or tails facing up. Once both coins have been placed, the players show their coins. If the coins match, Player A collects both coins, otherwise Player B collects them. The game matrix showing the payoff to Player A is

$$\begin{matrix} \square & 1 \\ \square & \\ \square & \\ \square & \\ \square & \end{matrix} \begin{matrix} \square & 1 \\ \square & \\ \square & \\ \square & \\ \square & \end{matrix} \quad (1)$$

If Player A uses the mixed strategy $x \cdot \text{Heads} + (1-x) \cdot \text{Tails}$, i.e. places his coin heads up x fraction of the time, and Player B uses the mixed strategy $y \cdot \text{Heads} + (1-y) \cdot \text{Tails}$, then the expected payoff of the game is

$$(1)xy + (\square 1)(1 \square x)y + (\square 1)x(1 \square y) + (1)(1 \square x)(1 \square y) = (1 \square 2x)(1 \square 2y). \quad (2)$$

Note that when A uses $x=1/2$, the outcome of the game is independent of B's strategy and vice versa. The game theoretic optimum for this game is $x^*=y^*=1/2$.

Now consider the evolution of two populations of players – N players of type A and M players of type B. Each player has a fixed mixed strategy $x_i, i=1 \dots N$ or $y_j, j=1 \dots M$. In one generation, each player's fitness is determined by some monotonic function of his total payoff after playing against a random sample of players of the other type. This is then used in some genetic drift and mutation (and possibly crossover) operators to determine the population of the next generation. Note that by the form of the payoff function in equation (2), as long as the A players maintain an average $\bar{x} = \sum x_i / N = 1/2$, all of the B players have the same expected payoff and so there is no selection pressure toward any B strategy! The corresponding argument holds for A players when $\bar{y} = 1/2$.

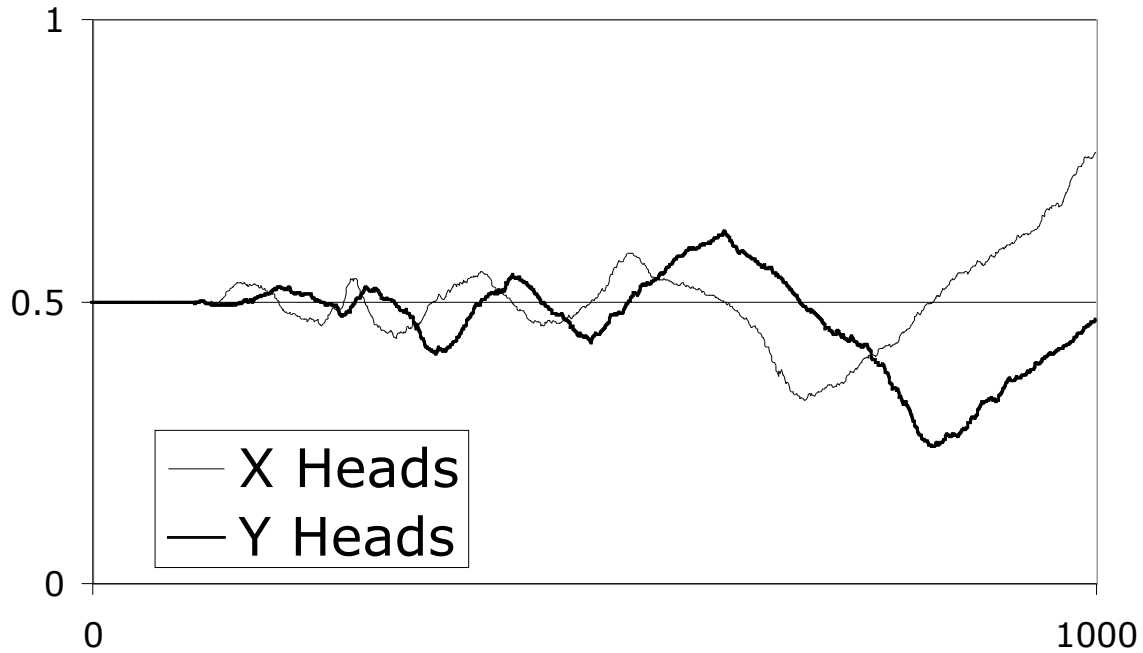


Figure 1: One hundred players on each of two teams playing Pennies. For the first 100 generations both teams have all players using the optimal strategies $x=y=1/2$. At the 100th generation, evolution begins and soon leads to sustained oscillations. Note how each curve turns around when the other curve crosses 0.5.

The result of a simulation experiments in this scenario is shown in Fig. 1. Both populations of players are started from their optimal strategies $x_i=y_j=1/2$, for all i and j . At each generation, a fitness is computed and a certain fraction of the population is removed with removal of the less fit being more likely. The removed players are replaced by random mutations of the survivors according to $x_{\text{new}}=x_{\text{old}}+\square$ and $y_{\text{new}}=y_{\text{old}}+\square$ where \square is a small random variable with mean 0. The drift term (\square) was kept equal to zero for the first 100 generations to show more dramatically the effect of the cycles that result. Initially both populations drift neutrally around a mean of 1/2. Once the mean of one

population has drifted significantly away, say $\bar{x} \gg 1/2$, the y 's start to feel a selection pressure. Notably, those y_j 's that are larger than $1/2$ are selected for and \bar{y} increases (recall that player B is trying to minimize the payoff). But once \bar{y} is larger than $1/2$, x feels a selection pressure to move toward $\bar{x} \ll 1/2$. The y 's do not feel this pressure until the x 's have moved past $1/2$. This starts the system cycling and the oscillations continue to grow until a regular pattern becomes established. We have performed extensive simulations confirming the phenomenon with a wide sample of selection and mutation operators in this and more complicated games including games whose optimal strategy is a mix of many pure strategies. The inherent noise in these systems induces a Neimark-Sacker bifurcation, the discrete analogue of a Hopf bifurcation. These systems are on the edge of a bifurcation in continuous time and become unstable by the addition of noise or by the discretization of time.

Conclusions and implications

The behavior described above is generic and should apply at least approximately in many biological, economic, and heuristic design situations. We begin by recalling the crucial ingredients in our model: (1) two populations of players repeatedly play a game of pure competition, (2) the fitness of each individual is determined by the aggregate winnings of that individual, (3) the strategies for one generation are fixed for each individual, and (4) the next generation "chooses" its strategies as a diffusion process originating dominantly from the more fit individuals in the previous generation.

The robustness and applicability of this process as an approximate model for many biological and economic situations result from relaxing the strict game-theoretic and genetic algorithmic aspects of the process. Pure competition could be relaxed to nearly pure competition. Generation can be reinterpreted to a market period. One important aspect of our implementation cannot be relaxed too much: the players cannot act like the rational all knowing agents of traditional game theory. Notably, the evolution of their strategies cannot use very long and accurate historical data sets. The requirement that the strategy choices be based only on current fitness can probably be relaxed slightly but the time horizon must stay small compared to one cycle. One could therefore call this a model of the "poorly informed agent".

Let us consider the economic setting first. The poorly informed agent model applies provided choosing the next strategies is done with limited historical information or foresight or processing power. Recent evidence (Bechara et al. 1997, Stewart 2002) indicates that many human decisions, including economic ones, are based on a gut feel of the estimated probabilities rather than elaborate forecasting. This fits our assumptions very nicely and predicts oscillatory behavior in these areas.

In other economic contexts, one "generation" can represent a year or other natural period for a consumer or producer to stick to one particular blend of (stores to shop in) or (crops to plant) or (advertising channels to use). The selection is naturally implemented by Adam Smith's invisible hand; the less successful shoppers or farmers or marketers are

likely to try the next year's fashion which is likely to resemble the behavior patterns of the more successful players.

In the biological setting, a strategy can represent producing a particular blend of toxins to deter competitors or predators. These levels of toxins are known to ebb and grow as the predators manage to produce antitoxins which show similar cycles (Maynard-Smith, 1982; Haynes and Sen, 1996; Czaran, et al., 2002). For biological applications it is particularly important that the phenomenon is robust with respect to making the reproductive times different for the two populations of players. Such robustness is in fact guaranteed. If one population reproduces more often, such extra reproductive cycles merely effect the rate at which the often reproducing population can move toward the optimal strategy to take advantage of any suboptimality in the other population. With appropriate selection and mutation parameters for the game of Pennies discussed above, we have seen one population follow what looks very much like a square wave with the mean strategy very close to 1 followed by the mean strategy very close to 0. The more often reproducing population will oscillate with larger amplitude than the slower population but oscillation is assured by the structure of the game.

In economic, biological or social contexts, the fact that the predicted oscillations are centered on the game theoretic optimum also leads to the following conclusion: the average aggregate payoff to both teams is the same as if all players on both sides always stuck to the optimal strategy. While this conclusion probably has many real-world applications, at the very least it says that equilibrium models that ignore dynamics

predict the correct average payoff. The diversity of strategies followed in the population is higher with the oscillations and this could well prove of evolutionary benefit in, for example, games with evolving payoffs.

Another set of implications of our results concerns the design of heuristics in learning systems. Many of the heuristics and learning rules currently implemented, for example evolutionary programming, work to coevolve optimal strategies. The current argument for the inherent instability of these attempts suggests that only one set of decisions be evolved at a time.

In real-life we play many games but the rules and payoffs are continually shifting and we see only a limited sample of outcomes. In many of these games, a model of current–performance based evolution, as treated here, is likely to describe the phenomena better than a rationally–selected evolution model. Cycles are everywhere and the cyclic nature of many systems (fashions, unstable cobwebs, state mandated teaching curricula, predator-prey adaptations) can be understood, at least partly, by the considerations of this paper. Specifically, our thesis is that they usually result from some game with a mixed optimal strategy. An advantage from the other player not pursuing his optimal strategy causes the first player to shift away from an all-bets-hedged position in turn exposing himself to the other player taking advantage, and so on. We hope that this paper can pave the way to recognizing such processes. There is more to iterated games than convergence to the Nash equilibrium.

Acknowledgements

We thank R. Whitney, A. Salamon, J. Mahaffy, A. Palacios, and the participants in the SDSU mathematical biology seminar for helpful discussions.

References:

Akiyama, E. and Kaneko, K. (1995) "Evolution of cooperation, differentiation, complexity, and diversity in an iterated three-person game." *Artificial Life* **2**, 293-304.

Axelrod, R. M. (1987). "The evolution of strategies in the iterated prisoner's dilemma," in *Genetic Algorithms and Simulated Annealing* (L. Davis, ed.) pages 32-41. Los Altos, Calif: Morgan Kaufmann.

Bechara A, Damasio H, Tranel D, Damasio AR (1997) "Deciding Advantageously Before Knowing the Advantageous Strategy." *Science* **275**:1293-1295

Boyd, R and Lorberbaum, J.P. (1987). "No pure strategy is evolutionarily stable in the repeated prisoner's dilemma game". *Nature*, **327**, 58-59.

Czaran, T., Hoekstra, R.F. and Pagie, L. (2002). "Chemical warfare between microbes promotes biodiversity," *PNAS*. **99**, 786-790.

Darwen, P. J. and Yao, X. (1995) "On evolving robust strategies for iterated prisoner's dilemma," in *Progress in Evolutionary Computation, Lecture Notes in Artificial Intelligence*, **956**, 276-292 (X. Yao, ed.), Heidelberg, Germany: Springer-Verlag.

Fogel, D.B. (1993). "Evolving Behaviors in the Iterated Prisoner's Dilemma," *Evolutionary Computation*, **1**, 77-97.

Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*, Reading, MA: Addison-Wesley.

Haupt, R., Haupt, S. (1998). *Practical Genetic Algorithms*, New York, NY: John Wiley & Sons.

Haynes, T. and Sen, S. (1996) "Evolving behavioral strategies in predators and prey," in *Adaptation and Learning in Multiagent Systems*, (G. Weiß and S. Sen, eds.) 113-126. Berlin: Springer Verlag.

Holland, J. (1975). *Adaptation in natural and artificial systems*. Ann Arbor: University of Michigan Press.

Koza, J. (1992). "Genetic evolution and co-evolution of game strategies" in *Proceedings of the international conference on game theory and its applications*, Stony Brook, New York.

Maynard-Smith, I.J. (1982). *Evolution and the Theory of Games*. Cambridge, MA: Cambridge University Press.

Rosin, C. and Belew, R. (1996). "New Methods for Competitive Co-evolution," *Evolutionary Computation* **5**, 1. <http://citeseer.nj.nec.com/rosin96new.html>

Rovatsos, M. and Lind, J. (1999) "Learning Cooperation in Repeated Games," in *Proceedings of the Workshop on Agents learning about, from and with other Agents* (IJCAI-99), Stockholm, Sweden (J. M. Vidal ed.).
<http://citeseer.nj.nec.com/rovatsos99learning.html>

Stewart, Thomas A. "How to Think With Your Gut" (2002) *Business 2.0*, November
<http://www.business2.com/articles/mag/0,1640,44584%7C3,00.html>

Tizard, M. (2000). "Artificial Life Programming Project - Sorting Networks by Co-evolution" <http://www.tizard.freewire.co.uk/Academic/Sorting/Index.htm>

Vajda, S. (1960). *An Introduction to Linear Programming and the Theory of Games*, New York, NY: Wiley.

van Nimwegen, E. James P. Crutchfield and Mitchell, M. (1997). "Finite populations induce metastability in evolutionary search," *Physics Letters A*, **229**, 144-150.

Yao, X. and Darwen, P. (1994) "An experimental study of N-person iterated prisoner's dilemma games," *Informatica*, **18**, 435-450.

<http://citeseer.nj.nec.com/yao94experimental.html>