

The efficiency of adapting aspiration levels

Martin Posch¹, Alexander Pichler² and Karl Sigmund^{2,3}

¹Universität Wien, Institut für Medizinische Statistik, Schwarzschanerstrasse 17, 1090 Wien, Austria

²Universität Wien, Institut für Mathematik, Strudlhofgasse 4, 1090 Wien, Austria

³International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria

Win-stay, lose-shift strategies in repeated games are based on an aspiration level. A move is repeated if and only if the outcome, in the previous round, was satisficing in the sense that the pay-off was at least as high as the aspiration level. We investigate the conditions under which adaptive mechanisms acting on the aspiration level (selection, for instance, or learning) can lead to an efficient outcome; in other words, when can satisficing become optimizing? Analytical results for 2×2 games are presented. They suggest that in a large variety of social interactions, self-centred rules (based uniquely on one's own pay-off) cannot suffice.

Keywords: games; satisficing; learning rules; natural selection

1. INTRODUCTION

In a game theory without rationality (see Rapoport 1984), players are not assumed to be able to fully understand the situation in which they are engaged. Their moves are based on knee-jerk rules rather than on strategic analysis. Possibly the simplest of such rules is the win-stay, lose-shift principle, which consists of repeating an action if it proved successful, and in switching to another action if not. Suppose that we were playing a machine with two levers, one resulting in a positive, the other in a negative outcome. The win-stay, lose-shift principle would result in our repeating the action with the positive outcome; if we erroneously tried the wrong action, we would switch back, in the next round, to the right action. Many experiments have shown that such a behaviour, or some approximation of it, is widespread among human and animal actors. Interestingly, this crudest form of a learning rule works even in situations involving several agents, as in the so-called minimal social situation (Colman 1995).

The win-stay, lose-shift principle was originally formulated by Thorndike:

'Of several responses made to the same situation, those which are accompanied or closely followed by satisfaction are more firmly connected with the situation; those which are accompanied or closely followed by discomfort have their connection with the situation weakened.' (Thorndike 1911, p. 244)

The wide range of validity of this principle was soon recognized (see for example, Hoppe 1931; Rescorla & Wagner 1972). In the hands of H. Simon, satisfaction-seeking behaviour became a leading contender for explaining social and economic decision making (see Simon 1955, 1957, 1962; Winter 1971; Radner 1975). A considerable amount of empirical evidence suggests that the behaviour of individuals and firms aims at satisficing, rather than optimizing.

But when do we feel satisfied? In certain situations (as when foraging for food, or for sex) our body knows. In other situations, we have to find out. We may feel pleased if we pulled a lever which delivers one dollar, but not if we are told that the alternative would have delivered ten. In such a situation, we must learn what to aim for, whereas in the foraging case our germ line has done the learning already and the result is encoded in the genome. Natural selection operating in a population, or a learning rule based on individual trial and error, can cause an adaptation of the aspiration level.

It is easy to see how selection, or learning rules, lead to an optimal aspiration level when playing against nature. We are interested in exploring how adaptation works when playing against other players. In the repeated prisoner's dilemma game, for instance, a strategy called PAVLOV does very well (see Kelley *et al.* 1962; Colman 1995; Kraines & Kraines 1988; Nowak & Sigmund 1993). PAVLOV is a win-stay, lose-shift rule with an aspiration level lying somewhere between the two highest and the two lowest pay-offs. Is there any reason to assume that selection or learning will adapt the aspiration level precisely to this interval? How would such adaptive mechanisms fare in other games? We will assume that our players are 'blind robots' without any knowledge of the structure of the iterated game, except that they have two options. They need not even be aware of the existence of another player. Their only information is the pay-off which they obtain in each round.

In §2, we shall briefly discuss some mechanisms for adapting the aspiration level, studying first the action of selection, and then two particularly simple learning rules, which are extremal cases of convex updating of the aspiration level, called YESTERDAY and FARAWAY. In §3–5, we turn to the simplest games, symmetrical games between two players having two strategies each. We examine whether adaptive mechanisms lead to an efficient outcome for such 2×2 games. This is one aspect of a larger question, namely: when is satisficing optimizing?

In this paper, our approach will be based on analytical methods. We restrict our attention to deterministic win-stay, lose-shift strategies based on switching to the alternative option if, and only if, the pay-off from the previous round falls below the aspiration level. (In Thorndike's formulation, win-stay, lose-shift is a stochastic rule: the difference between aspiration level and actual pay-off only affects the propensity to switch.) For a simulation-based exploration of win-stay, lose-shift strategies with longer memory sizes we refer to Posch (1999).

2. GAMES AGAINST NATURE

Consider a two-armed bandit. Pulling one lever yields pay-off R , pulling the other yields pay-off P , with $P < R$. Let a be the aspiration level of a player. The player will repeat the former action if the pay-off was at least a , and switch to the other action otherwise. With some probability $\epsilon > 0$ this action is misimplemented. For simplicity, we shall only consider the limiting case $\epsilon \rightarrow 0$ (that is, we compute the outcome for given $\epsilon > 0$ and then let ϵ converge to zero). We assume that the game consists of a large number of rounds, and that the pay-off for the repeated game is given by the limit-in-the-mean (LIM) of the pay-off per round (i.e. $\lim(p_1 + \dots + p_N)/N$ for $N \rightarrow \infty$, where p_n is the pay-off in round n). If $a > R$, the player will switch after every round, and obtain as LIM pay-off $(R + P)/2$. If $a \leq P$, the player will always be satisfied, switch only by mistake, and then repeat the new action until the next mistake occurs. Again the LIM pay-off is $(R + P)/2$. For $P < a \leq R$, the player will always pull the R -lever, except by mistake; after an erroneous P , the player will switch back to R . The LIM pay-off is R .

How does selection act on the frequencies x_1, x_2 and x_3 of the three strategies corresponding to the intervals $]-\infty, P]$, $]P, R]$ and $]R, +\infty[$ of possible aspiration levels? We shall assume that pay-off is converted into reproductive fitness, and that like begets like. This yields the replicator equation

$$\dot{x}_i = x_i(f_i - \bar{f}), \tag{1}$$

where f_i is the LIM pay-off for strategy i and $\bar{f} = \sum x_k f_k$ is the average LIM pay-off in the population (see Hofbauer & Sigmund 1998). The dynamics on the corresponding unit simplex S_3 lead to the extinction of the 'wrong' aspiration levels: x_2 converges to unity. In this sense, selection yields an aspiration level a in $]P, R]$.

What about learning? Conceivably the simplest way in which experience can affect a player's aspiration level consists in convex updating, by taking into account the pay-off obtained in the previous round. More precisely, if a_n is the aspiration level and p_n the pay-off in the n th round, then $a_n = (1 - \alpha)a_{n-1} + \alpha p_{n-1}$ for some fixed $\alpha \in]0, 1[$. If the aspiration level is initially higher than R , then the player will restlessly switch between the two possible actions, and a_n will steadily decrease until it is lower than R . If, however, a_n is lower than P , then the player will repeat the previous action. If this action happens to yield R , the aspiration level will soon be between R and P . If the action yields P , then a_n approaches P from below. A mistake in implementation will eventually bring it into the 'right' interval. Once

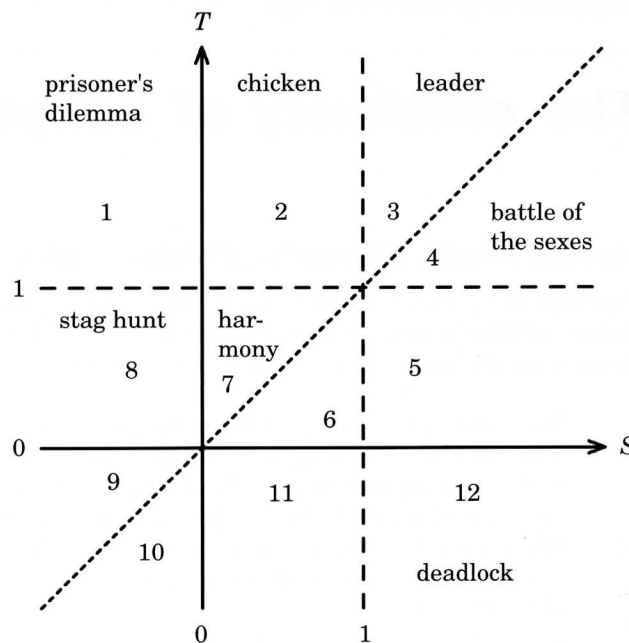


Figure 1. A partitioning of the (S, T) plane which displays the 12 symmetrical 2×2 games.

there, it will converge towards R from below. An eventual mistake in implementation happening now will not cause a_n to leave the interval $]P, R]$ and will immediately be corrected.

When players play each other (rather than a two-armed bandit), convex updating can lead to complex outcomes. We shall therefore restrict attention to two updating rules which represent two instructive extremal cases. With YESTERDAY, $\alpha = 1$, i.e. a_n is just p_{n-1} , the pay-off obtained in round $n - 1$. Even if a player starts with the P -lever, the first mistake will lead to the R -lever. The player then stays with this option: any further mistake will immediately be corrected.

FARAWAY is the opposite case, in some sense. Of course $\alpha = 0$ means no updating at all, which is uninteresting. Instead of this, we shall assume that the aspiration level is slowly but continuously modified towards the long-run average. This means that if the aspiration level is in $]-\infty, P]$ or $]R, +\infty[$, it steadily moves towards $(R + P)/2$ and eventually enters the interval $]P, R]$. Once there, it converges towards R . The direction of change defines dynamics leading asymptotically towards R , which is just 'right'.

3. 2×2 GAMES

The simplest non-trivial games involve two players with two options each, which we call C and D. We shall assume that the game is symmetrical, i.e. that the two players are interchangeable. The pay-off matrix is

$$\begin{pmatrix} R & S \\ T & P \end{pmatrix}, \tag{2}$$

i.e. R is the pay-off for using C against a player also using C, S for using C against D, etc. We consider only the generic situation where the four pay-off values are

