Learning Aspiration in Repeated Games*

In-Koo Cho†
and
Akihiko Matsui‡

May 12, 2003

Abstract

We study the behavior of boundedly rational agents who play an infinitely repeated symmetric 2×2 game according to a simple rule of thumb: each agent continues to play the same action if the action generates one period payoff that exceeds the aspiration level, which summarizes a history according to the average of the past payoffs, and switches to the other action with a positive probability otherwise. By applying the stochastic approximation technique (Kushner and Yin (1997)), we characterize the asymptotic outcomes and patterns of behavior for the class of symmetric 2×2 games, thus significantly extending results from the previous studies (Karandikar, Mookherjee, Ray, and Vega-Redondo (1998), Oechssler (2001), Posch (2001)). In a coordination game, for example, the two players achieve cooperation with probability one in the long run, while in the battle of the sexes, the outcome converges to either one of the two Pareto efficient outcomes, depending on the history, provided that the two players' one-shot equilibrium payoffs are not too "unfair". In the prisoners' dilemma game, the players cooperate in the limit if and only if the gain from defecting against cooperation is modest.

Keywords. aspiration, bounded rationality, cooperation, mean dynamics, recursive learning, repeated games, satisficing behavior, stochastic approximation

*We are grateful for helpful conversations with Oyama Daitsuke, Joseph Hofbauer, Michihiro Kandori, Martin Posch, and especially, Dilip Mookherjee as well as seminar participants at University of Tokyo, University of Illinois, University of Wisconsin, University of Vienna and Boston University. All remaining errors are ours.

†Financial support from the National Science Foundation (SES-0004315) is gratefully acknowledged. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Department of Economics, University of Illinois, 1206 S. 6th Street, Champaign, IL 61820 USA. E-mail: inkoocho@uiuc.edu. Web site: www.cba.uiuc.edu/inkoocho .

‡Financial support from the Ministry of Education and Science of Japan is gratefully acknowledged. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Ministry. Faculty of Economics, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033 JAPAN. E-mail: amatsui@e.u-tokyo.ac.jp. Web site: www.e.u-tokyo.ac.jp/~amatsui .

1 Introduction

According to the theory of satisficing behavior (Simon (1987)), a decision maker *satisfices* rather than optimizes.[1] Instead of solving an optimization problem, a decision maker has an aspiration level, and searches for a better alternative until he finds the one whose payoff exceeds the aspiration level. Once he finds such an alternative, he quits his search and sticks to it. The present paper examines the behavior of the satisficing decision makers in 2×2 symmetric repeated games, and examines the dynamic patterns of outcomes generated by the repeated interactions between the two satisficing agents.

We assume that each player continues to play the same action if the action generates one period payoff that exceeds the aspiration level, which summarizes a history according to the average of the past payoffs, and switches to the other action with a positive probability otherwise.[2] Focusing on the dynamic process of the aspiration levels, we study the asymptotic properties of the average payoffs and the patterns of the behavior of the decision makers.

However, the dynamic process of the aspiration levels might induce many different asymptotic outcomes, some of which do not appear to be sensible. In order to identify the sensible outcomes, we perturb the dynamics by introducing a small probability of imperfection: with a small probability, the decision maker might choose an action other than the one dictated by the "behavior" rule.

Because of the perturbation, the aspiration and the outcome of the game are rendered stochastic. As the action influences the aspiration levels, which in turn determine the future actions to be played, the evolution of the aspiration level exhibits a *self-referential* feature (Marcet and Sargent (1989)). Formally, the dynamics of the aspiration levels can be described as a state-dependent Markov process where the transition matrix of action pairs is determined by the aspiration level. Consequently, the evolution of the aspiration level and the resulting outcome path become non-stationary and history dependent, which poses a considerable challenge.

This paper uses the stochastic approximation technique (e.g., Kushner and Yin (1997)) to investigate the asymptotic properties of the aspiration levels. Instead of directly investigating the non-stationary stochastic process, we approximate the sample paths of the process by a trajectory induced by the deterministic process, which is called the mean dynamics or the ordinary differential equation (ODE) associated with the original process. This approximation enables us to study the asymptotic properties, which would otherwise be obscured by the complex stochastic perturbations. In our model, the associated ODE has a particularly simple form, which helps us directly calculate the limit points for all 2×2 games and examine the stability of these points.

We can completely characterize the asymptotic properties of the aspiration levels and the behavior of the players for symmetric 2×2 games. In the coordination game, we

[1]The satisficing behavior, or the "win-stay, lose-shift" principle, is observed in animal behavior as well as in human behavior. Thorndike (1911) mentioned this type of behavior prior to Simon (1987).

[2]Gilboa and Schmeidler (2001, pp.147-148) states, "Realism means that the aspiration level is set closer to the best average performance so far experienced."

have an intuition that the repeated interaction should promote cooperation between the two players, which is what our model predicts. Although many other outcomes can be sustained as a limit of deterministic dynamics, the Pareto efficient coordinating outcome is the only stable outcome of the perturbed dynamics.

The technique of stochastic approximation can be equally applied to games that do not have a *unique* Pareto efficient individually rational outcome such as the battle of the sexes.[3] In the battle of the sexes, we demonstrate that the learning dynamics lead to either one of the two efficient outcomes, but which of the two limit outcomes is realized depends upon the history. A similar analysis applies to the game of chicken.

In the prisoners' dilemma game, our model predicts that cooperation is achieved asymptotically if and only if the gain from defection is "modest."[4] If the gain is sufficiently large, cooperation is not attained. We can still characterize the limit points of the aspiration levels, from which one can infer the asymptotic behavior of the players. In other words, the size of the gain from defection determines the stability of the cooperation outcome.

In the prisoners' dilemma, if one player defects while the other player cooperates, then the cooperating player (who is double-crossed) receives the worst possible payoff of the game. The double-crossed player immediately switches his action to punish the defector, as we expect from the tit-for-tat strategy. Once both parties move to the non-cooperative outcome, both players immediately realize how bad the non-cooperative outcome could be, and move back to cooperation. If the gain from defection is not too large, this cycle harms the original defector as well as the defected. Thus, cooperation sustains. On the other hand, if the gain from defection is sufficiently large, the defector gains from the above cycle, and therefore, temptation for deviation pushes the outcome away from cooperation, and the cycle of cheating, punishment and cooperation repeats itself. As a result, the aspiration level converges (roughly) to the points obtained as a convex combination of three outcomes: cooperation, defection against cooperation and the one shot Nash equilibrium.[5]

Recently, the satisficing behavior has drawn considerable attention as an alternative behavioral assumption in games.[6] Gilboa and Schmeidler (1995) develops an axiomatic foundation for the satisficing behavior, and Gilboa and Schmeidler (2000) examines a more elaborate behavior by assuming that the agent adjusts the aspiration level as he accumulates experiences.

Karandikar, Mookherjee, Ray, and Vega-Redondo (1998, KMRV), Sarin and Vahid (1999), Oechssler (2001), and Posch (2001) focus on the component games which have a unique Pareto efficient individually rational outcome.[7] In KMRV, for example, each player uses his own past payoffs to calculate the aspiration level, while in Oechssler (2001), each player uses the average of the two players' payoffs. Both papers show that in a

[3] The battle of the sexes has two such pure strategy outcomes. See (7.14).

[4] See (7.16). To be precise, the gain from defection is "modest" if $g < 1$.

[5] There are two of such points, depending upon who cheats first against the other player at the cooperation outcome.

[6] Winter (1971) is a remarkable early example of the application of the satisficing behavior.

[7] Napel (2003) applies the method of KMRV to the ultimatum game.

repeated game with a unique Pareto efficient individually rational point in pure strategies, cooperation emerges in the limit. We freely borrow the basic elements of our model such as the dynamics of the aspiration level and the behavior rule from these papers.

Posch, Pichler, and Sigmund (1999) is essentially a deterministic version of ours, and considers all symmetric 2×2 games as we do. Since the model is deterministic, Posch, Pichler, and Sigmund (1999) does not address the issues that are dealt with in the present paper. As a result, for example, perpetual miscoordination arises in the coordination games. Our analysis, therefore, shows how stochastic terms play a role, while following the setup of the previous works closely.

At the same time, we significantly improve upon the existing studies, offering a comprehensive analysis for a larger class of games. By using the stochastic approximation technique, we develop a unified method of analysis, and consequently, avoid various issues arising from details of the specific games.[8] Indeed, once we identify the basic dynamic structure, the rest of the analysis is almost routine.

The technique of stochastic approximation has been used in the recursive learning literature in macroeconomics (e.g., Marcet and Sargent (1989), Evans and Honkapohja (2001)) and recently in evolutionary models (e.g., Benäim and Weibull (2001) and Sandholm and Hofbauer (2002)). Benäim and Weibull (2001) applies the technique to study the relationship between stochastic dynamics and deterministic dynamics such as best response dynamics in evolutionary models. In a repeated game between two players, each player has significant influence on the outcome from the game, and the learning dynamics of the two players are intertwined. As a result, the evolution of the state space is considerably more complicated than in the existing models where the diverse behavior of the agents is often aggregated into a one dimensional state variable. To handle the dynamics of the state variable in the learning dynamics, the stochastic approximation technique appears to be superior to the small perturbation method used in KMRV, for example, because of simplicity and generality of analysis.

Our paper can be viewed as an equilibrium selection theory in the repeated games. A number of attempts have been made to mitigate this problem of multiplicity since the seminal paper of Aumann (1959). The main stream of these attempts is to choose equilibria that can be attained by the boundedly rational players. Depending upon how restrictive the bound on the rationality is, we can classify the general approaches into two groups.

The first is to maintain the optimizing behavior while limiting the computational capability of the decision maker but often endowing the agent with the ability to compute an equilibrium (e.g., Rubinstein (1986), Abreu and Rubinstein (1988), Aumann and Sorin (1989)). The second approach imposes a very tight bound on the computational capability of the decision maker by assuming adaptive and/or imitating behavior (e.g., Carmichael and McLeod (1997) and Vega-Redondo (1997)). The satisficing behavior falls in between

[8]For example, we do not have to introduce the kind of mutation on top of perturbation as was necessary in KMRV. Also, we do not encounter some of the problems that the previous works would encounter when one tries to study the games which have multiple individually rational pure strategy outcomes that are Pareto efficient such as the battle of the sexes.

these two formulations. In the equilibrium approach with optimizing agents, players have limited capacity to implement their strategies, they have unlimited capacity to read the opponent's strategy, or at least there is assumed to be a mechanism by which equilibrium is attained. On the other hand, the second, adaptive and/or imitating, behavior corresponds to the evolutionary approach. Players are not required to have a sufficient amount of computational capacity and can even be viewed as mindless genes who are subject to the survival of the fittest. Satisficing behavior has been developed as the theory of human behavior which is neither perfectly rational nor completely "mindless."[9]

The satisficing behavior (or case-based decision theory) has been criticized for lacking fruitful application. Indeed, in single-person decision making problems, the expected utility theory and the case-based decision theory are equivalent in the sense that one can be embedded in the other as proven by Matsui (2000). The present paper shows, however, that the two theories produce very different results in the two-player world. This is because the satisficing behavior does not require an initial belief to be formed.

The rest of the paper is organized as follows. Section 2 presents the basic model with the deterministic behavior rule. Section 3 considers an example of common interest games to illustrate what the learning dynamics can implement in the limit and how we should refine the limit point to select the sensible outcome. Section 4 formalizes the idea of perturbing the decision rule. Section 5 formally presents a model with perturbation. Section 6 explains the technique of stochastic approximation. Section 7 characterizes the stability points of the stochastic dynamics in 2×2 games. Section 8 analyzes a model in which the aspiration level is determined as a convex combination of the two players' payoffs in order to illustrate how our analysis applies to a larger class of models including Posch (2001) and Oechssler (2001). Section 9 concludes the paper.

2 Unperturbed model

Consider two players who play a 2×2 game

$$G = \langle \{C, D\}, \{C, D\}; u_1, u_2 \rangle$$

infinitely many times, where C and D are the actions available for each player and

$$u_i : \{C, D\} \times \{C, D\} \to \mathbb{R} \qquad i = 1, 2$$

is the payoff function of the one shot game. Let $s_t \in \{C, D\}^2$ be an outcome in period t. A history is a sequence of outcomes, and a repeated game strategy is a mapping from a history to an action $s_i \in \{C, D\}$. Let us assume that each individual player is sufficiently patient so that the long run cooperation at (C, C) can be sustained by some equilibrium of the infinitely repeated game.

[9] The difference between adaptive behavior and satisficing behavior might be just a matter of interpretation. However, we would like to draw the line in the sense that the latter has been developed as the model of human behavior that is not conceivable in genetic behavior.

Instead of a fully rational player, we model each player as a boundedly rational decision maker who summarizes a past history into a single state variable according to a simple rule of thumb, which we call the *aspiration level.* Let $a_{i,t}$ be the aspiration level of player i at the t-th round, which is calculated as the average of the past payoffs:

$$a_{i,t} = \frac{1}{t}\sum_{\tau=1}^{t} u_i(s_\tau).$$

Note that $a_{i,t}$ can be written recursively as

$$a_{i,t} = a_{i,t-1} + \gamma_t \left(u_i(s_t) - a_{i,t-1}\right) \tag{2.1}$$

where

$$\gamma_t = \frac{1}{t}.$$

Note that as player i summarizes a history into the average payoff in the past, he suppresses much information that could have been used for making a choice, such as the precise sequence of outcomes.[10]

The aspiration level $a_{i,t}$ is the level of payoff which player i expects from action $s_{i,t}$ in period t, based on his own past experience. Thus, if player i has entertained a sequence of high payoffs, it seems natural that player i continues to expect to receive a high payoff from his action, and vice versa. If today's payoff $u_i(s_t)$ from player i's action $s_{i,t} \in S_i = \{C, D\}$ exceeds aspiration level a_i^t, then player i takes s_i in the $t+1$st period. If $u_i(s_t) \leq a_i^t$ holds, then player i takes $s_i' \neq s_i$ in the next period. To describe the behavior rule rigorously, let $-s_i$ be the action other than $s_i \in \{C, D\}$:

$$\{-s_i\} = \{C, D\} \setminus \{s_i\}.$$

For all $i \in \{1, 2\}$,

$$s_{i,t+1} = \begin{cases} s_{i,t} & \text{if} \quad u_i(s_{1,t}, s_{2,t}) > a_{i,t} \\ (1-p)[-s_{i,t}] + p[s_{i,t}] & \text{if} \quad u_i(s_{1,t}, s_{2,t}) \leq a_{i,t}, \end{cases} \tag{2.2}$$

where $p \in (0, 1)$ represents inertia, and $(1-p)[-s_{i,t}] + p[s_{i,t}]$ is a mixed strategy over $s_{i,t}$ and $-s_{i,t}$ with $s_{i,t}$ being played with probability p.

We introduce the inertia to capture the idea of the switching cost. That is, once a player chooses an action, it would be more difficult to switch to another action than to play the same action in the next round (cf. KMRV). One can interpret p close to 0 as a minimal switching cost. Similarly, if p is close to 1, the switching cost is so large that the agent might not move to the alternative action.

The inertia will play an important role in breaking a seemingly unrealistic coordination failure of actions between the two players. Without inertia, players could end up

[10] The following analysis will be valid with some modification even if γ_t is constant across time. In that case, $\gamma_t = \gamma$ has to be sufficiently small, and the notion of convergence to be used will be the convergence in distribution as opposed to the strong convergence.

simultaneously switching back and forth between two off-diagonal action profiles in, say, a repeated coordination game, which will be shown in Section 3. No conclusion of the paper is sensitive to the specific tie breaking rule in (2.2). We shall exclude the mixed strategy for the moment.

Combining (2.1) and (2.2) for given initial aspiration levels $a_0 = (a_{1,0}, a_{2,0})$ and actions $s_0 = (s_{1,0}, s_{2,0})$, we can generate $\{a_t, s_t\}_{t=1}^{\infty}$. We are most interested in characterizing the asymptotic properties of $\{a_t, s_t\}_{t=1}^{\infty}$ in order to understand the asymptotic behavior of the players.

3 Example

Let us consider the following coordination game

$$\begin{array}{cc} & \begin{array}{cc} C & \quad D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} 1,\ 1 & -1,\ -1 \\ -1,\ -1 & 0,\ 0 \end{pmatrix} \end{array} \tag{3.3}$$

in order to illustrate the learning dynamics. Assume throughout this example that the initial aspiration level $a_{i,0}$ for player i is slightly less than 0, say -0.1.

If each player chooses C in the initial round, then the realized payoff for player i is 1, which is strictly larger than his aspiration level in the initial round -0.1. Thus, player i chooses C in the next round. Because the average payoff is less than the payoff from (C, C), each player keeps playing C so that the aspiration vector $a_t = (a_{1,t}, a_{2,t})$ converges to $(1, 1)$, which is the most sensible outcome from this game.

Perpetual coordination failure is avoided thanks to inertia. If it were not for inertia, (C, D) would be followed by (D, C) without fail provided that the aspiration level for each player is above -1; (D, C) and (C, D) would alternate, and no desirable outcome would be attained.

While we can sustain the Pareto efficient coordination outcome through the aspiration learning process, we can also sustain other outcomes, which are far less intuitive. Suppose that the initial action pair is (D, D). Then following behavior rule (2.2), the action pair stays at (D, D), while the aspiration level pair converges to $(0, 0)$ from below, but never exceeds it.

The convergence process is not stable. For example, suppose that each player experiments by playing a different action with a small probability $\rho > 0$, which could be also interpreted as the probability of making a mistake. Then with a small, but positive, probability $\rho > 0$, one of the two players switches his action to C to reach either (C, D) or (D, C), which is followed by (C, C) with a positive probability of $p(1-p)$, i.e., cooperation is achieved. Once (C, C) is realized, the two players maintain C since the realized payoff $u_i(C, C) = 1$ exceeds the aspiration level. The repetition of the inefficient coordinating outcome is not robust against a small probability of mistakes.

While informal, this example reveals the need for small perturbations to eliminate unintuitive outcomes. To this end, we introduce a small probability of choosing a different action from what is instructed, which is the subject of the next section.

4 Approximation by a smooth function

Note that the behavior rule is represented by a step function of

$$u_i(s_t) - a_{i,t}.$$

If $u_i(s_t) - a_{i,t} > 0$, then player i chooses the same action in period $t+1$ as he did in period t. Otherwise, he switches to another action. The discontinuity occurs at the point where $u_i(s_t) - a_{i,t} = 0$.

We perturb this decision rule by introducing a small amount of imperfection in switching decisions. If $|u_i(s_t) - a_{i,t}|$ is large, then it is very clear that the present action exceeds or falls short of the aspiration level expected from the action. Thus, there is little chance to get confused about switching the action. On the other hand, if $|u_i(s_t) - a_{i,t}|$ is close to 0, then the agent might not be so sure about whether to switch the action or not. Each player is suspicious about the possibility of imperfect observation, or a small observational error. Thus, when $|u_i(s_t) - a_{i,t}|$ is small, player i has a good reason to hesitate to accept the observation at its face value.

To formulate this idea, let $h_i : \mathbb{R} \to (0,1)$ be a continuously differentiable weakly increasing function satisfying the following properties:

$$\lim_{x\to\infty} h_i(x) = 1, \quad \text{and} \quad \lim_{x\to-\infty} h_i(x) = p. \tag{4.4}$$

We call such h_i a *sigmoid* function.

Let $h_i\left(u_i(s_t) - a_{i,t}\right)$ be the probability that player i chooses the same action in $t+1$ as in t. Because h_i is strictly increasing, player i will choose the same action in period $t+1$ with a large probability if $u_i(s_t) - a_{i,t} > 0$ is very large. Similarly, if the present payoff $u_i(s_t)$ falls "far" below aspiration level $a_{i,t}$, then player i switches to another action with a large probability, albeit subject to inertia.

It is not necessary that $h_i(x)$ be greater than p, and *strictly* smaller than 1 over the entire domain of x. We impose this condition only to simplify the analysis. The same analysis applies as long as there exists $\epsilon > 0$ such that $0 < h_i(x) < 1$ for all $x \in N_\epsilon(0)$, where $N_\epsilon(0)$ is the ϵ-neighborhood of 0. On the other hand, it is essential that

$$0 < h_i(0) < 1$$

which forces player i to experiment with a positive frequency as long as $u_i(s_t) - a_{i,t}$ is close to 0, *regardless* of its sign.[11]

[11] KMRV used a different sigmoid function $\tilde{h}_i$ satisfying $\tilde{h}_i(0) = 1$ so that for $\forall x \geq 0$, $\tilde{h}_i(x) = 1$. That is, as long as $u_i(s_t) - a_{i,t} \geq 0$, $s_{i,t+1} = s_{i,t}$ with probability one. This seemingly minor difference leads to profoundly different asymptotic properties of the learning dynamics in the prisoner's dilemma game when the gain from deviation against cooperation is large, i.e., $g > 1$. While we predict that cooperation outcome is not stable, KMRV predicts the cooperation is the only limit point of the learning dynamics. One can view our model as the "perturbed" version of KMRV in the sense that player i's behavior rule is subject to a tremble regardless of the sign of $u_i(s_t) - a_{i,t}$.

5 Perturbed model

Let us consider the sigmoid function h_i defined as (4.4). If the present payoff falls below the aspiration level, player i can stick to the present action with probability p. For a pair h and h' of increasing real valued functions over $\mathbb{R}$, let $d(h, h')$ be the Hausdorff metric of the graphs of h and h' in $\mathbb{R}^2$. In particular, if h' is a step function and h is a sigmoid function, $d(h, h') \to 0$ implies that $h(x) - h'(x) \to 0$ pointwise except at discontinuous points. Define

$$\overline{h}^p(x) = p + (1-p)\mathbf{1}_{x \geq 0}$$

where $\mathbf{1}_W$ is the characteristic function of event W: $\mathbf{1}_W = 1$ if the state is in W, and 0 otherwise. We omit p to write $\overline{h}$ whenever there is no risk of confusion. We are interested in the dynamics in the limit of the sigmoid functions which converge to $\overline{h}$.

We arrange the elements of S according to the following order:

$$(C, C), (C, D), (D, C), (D, D)$$

and let $\sigma \in \Delta^4$ (which is a row vector in the unit simplex of $\mathbb{R}^4$) be the probability distribution over S. Let $P(a) = (p_{s's}(a))$ be the transition matrix induced by h_i where

$$\begin{aligned} p_{s's}(a) = \quad & \left[h_1(u_1(s') - a_1)\mathbf{1}_{s_1 = s_1'} + (1 - h_1(u_1(s') - a_1))\left(1 - \mathbf{1}_{s_1 = s_1'}\right)\right] \\ & \times \left[h_2(u_2(s') - a_2)\mathbf{1}_{s_2 = s_2'} + (1 - h_2(u_2(s') - a_2))\left(1 - \mathbf{1}_{s_2 = s_2'}\right)\right] \end{aligned}$$

is the element in the s'-th row and s-th column. Note that every component of $P(a)$ is a strictly positive differentiable function of $a = (a_1, a_2)$. Similarly, define $\overline{P}(a) = (\overline{p}_{s's}(a))$ induced by $\overline{h}$ in which

$$\begin{aligned} \overline{p}_{s's}(a) = \quad & \left[\overline{h}_1(u_1(s') - a_1)\mathbf{1}_{s_1 = s_1'} + \left(1 - \overline{h}_1(u_1(s') - a_1)\right)\left(1 - \mathbf{1}_{s_1 = s_1'}\right)\right] \\ & \times \left[\overline{h}_2(u_2(s') - a_2)\mathbf{1}_{s_2 = s_2'} + \left(1 - \overline{h}_2(u_2(s') - a_2)\right)\left(1 - \mathbf{1}_{s_2 = s_2'}\right)\right]. \end{aligned}$$

Note that

$$\sigma P(a)$$

represents the action realized in the present round, conditioned on the previous pair of actions, σ, and aspiration vector a. Let u be a 4×2 matrix obtained by arranging the component game payoff:

$$u = \begin{bmatrix} u_1(C, C) & u_2(C, C) \\ u_1(C, D) & u_2(C, D) \\ u_1(D, C) & u_2(D, C) \\ u_1(D, D) & u_2(D, D) \end{bmatrix}.$$

Given an action distribution σ, $\sigma \cdot u$ represents the expected payoff vector.

It would be useful to review some of the properties of $P(a)$.

Lemma 5.1 [1] *For any a, $P(a)$ is irreducible and aperiodic so that it has a unique invariant distribution σ_a^* satisfying*

$$\sigma_a^* = \sigma_a^* P(a)$$

and for any initial condition σ_0,

$$\lim_{t\to\infty} \sigma_0 P^t(a) = \sigma_a^*.$$

[2] *σ_a^* is continuous with respect to a.*

[3] *If $\overline{P}(a)$ has a unique invariant distribution $\overline{\sigma}_a$, then σ_a^* converges to $\overline{\sigma}_a$ as $d(h_i, \overline{h}) \to 0$.*

The first property allows us to "represent" $P(a)$ by its invariant distribution, which significantly simplifies the characterization of the learning dynamics. Because the invariant distribution changes continuously with respect to the transition matrix, we often "approximate" transition matrix $P(a)$ by $\overline{P}(a)$, which has a significantly simpler structure than $P(a)$.

6 Stochastic approximation

To investigate the dynamics of a_t, one can examine the sample path induced by s_t and a_t in discrete time. Instead, we examine continuous time approximations of s_t and a_t obtained by proper interpolations. This detour illuminates the key analytic method, which is also used to examine the stochastic dynamics. To make this paper self-contained, we briefly summarize the classic results from the stochastic approximation, needed for our analysis, mostly from Kushner and Yin (1997).

Recall (2.1). For $K = 1, 2, \ldots$, define

$$t_K = \sum_{t=1}^{K} \gamma_t$$

as the total amount of (fictitious) time that is needed to play K rounds. We view $u(s_t) - a_{t-1}$ as the amount of change of the aspiration level during 1 unit of (fictitious) time, and γ_t as the amount of (fictitious) time assigned to this round t. Thus, the net change that can be made to the aspiration level during the t-th round that lasts γ_t units of (fictitious) time is $\gamma_t \left(u(s_t) - a_{t-1}\right)$. For $\forall \tau > 0$, there is a unique K such that

$$t_{K-1} \le \tau < t_K.$$

Define $m(\tau) = K$ as the first round that passes τ time. Let $\overline{a}(\tau)$ be the continuous time process obtained from a_t through linear interpolation. For $\tau \in [t_{K-1}, t_K)$,

$$\overline{a}(\tau) = \frac{t_K - \tau}{\gamma_K} a_{K-1} + \frac{\tau - t_{K-1}}{\gamma_K} a_K.$$

Because the action space is discrete, we construct the continuous time counterpart $\overline{s}(\tau)$ of s_t through constant (not linear) interpolation:

$$\overline{s}(\tau) = s_{K-1} \quad \forall \tau \in [t_{K-1}, t_K).$$

To understand the asymptotic properties of a_t, it suffices to examine the trajectory of $\overline{a}(\tau)$. For a small $\tau > 0$, the trajectory of $\overline{a}(\tau)$ is jagged. However, as $\tau \to \infty$, the gain function γ_t becomes smaller so that the trajectory of $\overline{a}(\tau)$ becomes "smoother." To formalize this intuition, let us fix $\tau > 0$ and consider

$$\overline{a}(t_K + \tau) - \overline{a}(t_K). \tag{6.5}$$

We then rescale the time by setting t_K equal to 0, which is called the *left shift of time*. Define

$$\overline{a}^K(\tau) = \overline{a}(t_K + \tau).$$

Then, (6.5) can be written as

$$\overline{a}^K(\tau) - \overline{a}^K(0). \tag{6.6}$$

If the behavior rule is deterministic, σ is a degenerate probability distribution that is concentrated on a particular pair of actions. We use σ in place of the pair of actions.

Recall that

$$\sigma_{t-1} P(a_{t-1}) = \sigma_t.$$

Define

$$\delta M_t = u(s_t) - \sigma_t u$$

which is a martingale difference:

$$E_t \delta M_t = 0 \qquad \forall t \geq 1$$

where E_t is the expectation conditioned on information available at period t, because s_t is distributed according to σ_t. Clearly,

$$|\delta M_t| \leq \max_{s,s',i} |u_i(s) - u_i(s')|.$$

To simplify notation, let us define $O(h)$ as a function satisfying

$$\lim_{h \to 0} O(h) = 0.$$

For a fixed $\tau > 0$,

$$\begin{aligned}
& \overline{a}^K(\tau) - \overline{a}^K(0) \\
= & \sum_{t=K+1}^{m(t_K+\tau)-1} \gamma_t (u(s_t) - a_{t-1}) + O(\gamma_{m(t_K+\tau)}) \\
= & \sum_{t=K+1}^{m(t_K+\tau)-1} \gamma_t (\sigma_{t-1} P(a_{t-1}) u - a_{t-1}) + \sum_{t=K+1}^{m(t_K+\tau)-1} \gamma_t \delta M_t + O(\gamma_{m(t_K+\tau)}).
\end{aligned}$$

Note that for $\forall \tau > 0$, there exist $K', M > 0$ such that $\forall K \geq K'$,

$$\frac{\overline{a}^K(\tau) - \overline{a}^K(0)}{\tau} \leq M.$$

Define

$$\delta_{t-1} = a_{t-1} - \overline{a}^K(0)$$

for $t \geq K+1$. Note that

$$\sigma_{t-1} P(a_{t-1}) = \sigma_K P(a_K) P(a_{K+1}) \cdots P(a_{t-1}) = \sigma_K P(a_K) P(a_K + \delta_1) \cdots P(a_K + \delta_{t-1}).$$

Consider

$$b_t = a_K P(a_K) \cdots P(a_K + \delta_t).$$

Since $\{b_t\}$ and $\{a_t\}$ are contained in compact sets, each sequence has a convergent subsequence. After renumbering the sequences, we have $b_t \to b^*$ and $\delta_t \to \delta^*$ for some b^*, δ^*. Since $P(a)$ is a continuous function of a,

$$b^* P(a_K + \delta^*) = b^*.$$

Given $\tau > 0$,

$$\|b^* - \sigma^*_{a_K}\| = O(\tau)$$

where

$$\sigma^*_{a_K} P(a_K) = \sigma^*_{a_K},$$

because $P(a)$ is uniformly continuous and has a unique invariant distribution for each a. Then, we have

$$\begin{aligned}
&\overline{a}^K(\tau) - \overline{a}^K(0) \\
= &\sum_{t=K+1}^{m(t_K+\tau)-1} \gamma_t \left(\sigma_K P(a_K) \cdots P(a_{t-1}) u - a_{t-1}\right) + \sum_{t=K+1}^{m(t_K+\tau)-1} \gamma_t \delta M_t + O(\gamma_{m(t_K+\tau)}) \\
= &\sum_{t=K+1}^{m(t_K+\tau)-1} \gamma_t \left(\sigma_K P^{t-K}(a_K) u - a_{t-1}\right) + \sum_{t=K+1}^{m(t_K+\tau)-1} \gamma_t \delta M_t + O(\gamma_{m(t_K+\tau)}) \\
= &\sum_{t=K+1}^{m(t_K+\tau)-1} \gamma_t \left(\sigma^*_{a_K} u - a_{t-1}\right) + \sum_{t=K+1}^{m(t_K+\tau)-1} \gamma_t \delta M_t + O(\gamma_{m(t_K+\tau)}) + \tau O(1/K)
\end{aligned}$$

where the last equality follows from the fact that $\sigma^*_{a_K}$ is the unique invariant distribution of $P(a_K)$. Since the invariant distribution σ^*_a is a continuous function of a,

$$|\sigma^*_{a_{t-1}} - \sigma^*_{a_K}| \leq O(\tau) \qquad \forall t \in \{K+1, \ldots, m(t_K+\tau)-1\}.$$

Thus, we have

$$\begin{aligned}
&\overline{a}^K(\tau) - \overline{a}^K(0) \\
= &\sum_{t=K+1}^{m(t_K+\tau)-1} \gamma_t \left(\sigma^*_{a_{t-1}} u - a_{t-1}\right) + \sum_{t=K+1}^{m(t_K+\tau)-1} \gamma_t \delta M_t + O(\gamma_{m(t_K+\tau)}) + \tau O(1/K) + \tau O(\tau).
\end{aligned}$$

Note that the first term approximates the Riemann integration:

$$\sum_{t=K+1}^{m(t_K+\tau)-1} \gamma_t \left(\sigma^*_{a_K} u - a_{t-1}\right) = \int_0^\tau \sigma^*_a u - a(\tau')d\tau' + \tau O(1/K)$$

where $O(1/K)$ is the interpolation error arising from the Riemann integration. Therefore,

$$\begin{aligned}
\overline{a}^K(\tau) - \overline{a}^K(0) \quad = \quad & \int_0^\tau \sigma^*_a u - a(\tau')d\tau' + \sum_{t=K+1}^{m(t_K+\tau)-1} \gamma_t \delta M_t \\
& + O(\gamma_{m(t_K+\tau)}) + \tau O(1/K) + \tau O(\tau) + \tau O(1/K). \qquad (6.7)
\end{aligned}$$

Note that except for the first term, the remaining terms on the right hand side are "small" if $\tau > 0$ is small and K is large. Thus, it would be reasonable to expect that the right hand side is "dominated" by the first term, or more formally, the trajectory of the right hand side is close to the trajectory induced by the first term.

Let us consider an ordinary differential equation (ODE)

$$a(\tau) - a(0) = \int_0^\tau \sigma^*_{a(\tau')} u - a(\tau')d\tau'$$

or equivalently,

$$\frac{da}{d\tau} = \sigma^*_{a(\tau)} u - a(\tau) \qquad (6.8)$$

for a given initial condition $a(0)$, where

$$\sigma^*_a = \sigma^*_a P(a).$$

To simplify notation, we often write the right hand side of ODE as Ψ:

$$\dot{a} \equiv \Psi(a) = \left[\Psi_1(a), \Psi_2(a)\right]. \qquad (6.9)$$

We call (6.9) the associated ODE, or the mean dynamics of a_t.

In fact, this approximation result can be proved for any $\tau > 0$ as $K \to \infty$, which is the fundamental result first proved by Ljung (1977) and Kushner and Clark (1978).

Lemma 6.1 *For any $\tau > 0$,*

$$\lim_{K\to\infty} \overline{a}^K(\tau) - \overline{a}^K(0) - \int_0^\tau \Psi(\tau')d\tau' = 0$$

with probability one.

Proof. See Appendix A.

Lemma 6.1 is key to approximating a jagged stochastic process by a differentiable deterministic path induced by the associated mean dynamics. Once we find the mean dynamics, what remains is to characterize its solution. Therefore, the main task would be to calculate the deterministic process given by (6.9).

Definition 6.2 *$a^* \in \mathbb{R}^2$ is a stationary state if*

$$\Psi(a^*) = 0.$$

Stability of stationary states is central for our analysis.

Definition 6.3 *A stationary state a^* is locally stable (in the sense of Lyapunov) if there exists $\overline{\rho} > 0$ such that if $a(0) \in N_{\overline{\rho}}(a^*)$, then for all $\rho \in (0, \overline{\rho})$, there exists $T > 0$ such that for all $\tau \geq T$, $a(\tau) \in N_\rho(a^*)$, where $N_\rho(\cdot)$ is the open ball with radius ρ.*

If a^* is not locally stable (in the sense of Lyapunov), then we say a^* is not stable. If a^* is locally stable (in the sense of Lyapunov) where the entire space is the domain of attraction, then a^* is globally stable.

Although the class of games we analyze has no limit cycle in the system, one might want to generalize the notion to a set-valued one. In that case, we may replace the set of locally stable states by the following notion of set-valued stability.

Definition 6.4 *We say that the set $\mathcal{K}$ is a locally stable set with respect to (6.9) if $\mathcal{K}$ is a minimal compact set with respect to the following properties. There exists $\overline{\rho} > 0$ such that for all $\rho \in (0, \overline{\rho})$ and for all $a(0) \in N_{\overline{\rho}}(\mathcal{K})$, there exists $T > 0$ such that for all $\tau \geq T$, $a(\tau) \in N_\rho(\mathcal{K})$.*

This condition says that if $a_t \in \mathcal{K}$, then the trajectory must converge to $\mathcal{K}$. This condition still admits the possibility that a_t oscillates between two different points in $\mathcal{K}$.

The next theorem, which is adopted from the standard convergence theorem in the stochastic approximation literature, is central for our investigation.

Theorem 6.5 *Suppose that $\mathcal{K}$ is the family of locally stable sets of (6.9) with the domain of attraction containing V. Then for all $a_0 \in \mathbb{R}^2$ and all s_0,*

$$a_t \to \mathcal{K}$$

with probability one.

Proof. See Appendix B.

By invoking Theorem 2 of Woodford (1990), we can also show that if a stationary state is not locally stable, then the asymptotic probability distribution of a_t assigns 0 probability to its neighborhood. Thus, the asymptotic probability distribution of a_t must be concentrated at the neighborhood of the stable states of (6.9).

In principle, the limit set $\mathcal{K}$ depends upon a pair of sigmoid functions $h = (h_1, h_2)$. Given h, let $\mathcal{K}_h$ be the limit set as selected in Theorem 6.5. In the following section, we are interested in the limit of $\mathcal{K}_h$ as h converges to $\overline{h}$.

7 Analysis

By Theorem 6.5, it suffices to analyze the stable states of (6.9) to characterize the asymptotic distribution of the aspiration levels induced by the learning dynamics. Although the complexity of the analysis varies substantially for different games, these games have the same underlying properties, some of which we shall see first.

Let V be the set of all feasible payoff vectors:

$$V = \left\{ \sum_{s \in S} \alpha_s u(s) \mid \alpha_s \geq 0, \sum_{s \in S} \alpha_s = 1 \right\}. \tag{7.10}$$

Since we consider the case of a small amount of perturbation, each component of the transition matrix $P(a)$ is approximated by $\overline{P}(a)$ induced by $\overline{h}$ except at thresholds. Recall that for a fixed p, $\overline{h}$ is completely determined by

$$(u_1(s) - a_1, u_2(s) - a_2).$$

Thus, it is convenient to partition V into

$$\mathcal{S} = \{\mathcal{S}_1, \ldots, \mathcal{S}_M\}$$

where for all $m = 1, \ldots, M$,

$$a, a' \in \mathcal{S}_m \text{ if and only if } \quad \forall s \in S,\ \forall i = 1, 2,\ a_i > u_i(s) \Rightarrow a'_i > u_i(s).$$

For any a and a' in the same quadrant $\mathcal{S}_m$, $\overline{P}(a) = \overline{P}(a') = \overline{P}_m$.

Take an arbitrary $\mathcal{S}_m$. Then calculate an invariant distribution $\overline{\sigma}^m = \overline{\sigma}^m \overline{P}_m$. The unperturbed dynamics in this quadrant are approximated by

$$\frac{da}{dt} = \overline{\sigma}^m(s) u - a. \tag{7.11}$$

Note that some component in $\overline{P}(a)$ might be zero, and therefore, $\overline{P}(a)$ may have multiple invariant distributions. On the other hand, thanks to the sigmoid function h_i, $P(a)$ is aperiodic and irreducible, and therefore, $P(a)$ has a unique invariant distribution.

If $\overline{P}(a)$ has a unique invariant distribution, then Lemma 5.1 [3] implies that the invariant distribution of $P(a)$ converges to that of $\overline{P}(a)$. If, on the other hand, $\overline{P}(a)$ has multiple invariant distributions, the upper hemi-continuity property of the invariant distribution holds: when $d(h_i, \overline{h})$ is small, the invariant distribution of $P(a)$ must be close to some, if not all, invariant distribution of $\overline{P}(a)$.

Because the mean dynamics are dictated by the invariant distribution of $P(a)$, we can replace $P(a)$ by $\overline{P}(a)$ (with some care when $\overline{P}(a)$ has multiple invariant distributions). Because the calculation of the invariant distribution is easy, and does not require any specific details of the component game, it is straightforward to analyze *any* 2×2 game by using the same method. In this paper, we examine only symmetric 2×2 games, while leaving other 2×2 games for interested readers.

7.1 Coordination games

In order to illuminate the basic idea of the analysis, we first reexamine the pure coordination game (3.3) and demonstrate that the sigmoid function and the inertia help the aspiration level converge to $u(C, C)$.

In (3.3), we have

$$V = \{(a_1, a_2) \mid -1 \leq a_1 = a_2 \leq 1\}.$$

We partition V into two pieces:

$$\mathcal{A} = \{(a_1, a_2) \mid -1 \leq a_1 = a_2 \leq 0\}$$

and

$$\mathcal{B} = \{(a_1, a_2) \mid 0 < a_1 = a_2 \leq 1\},$$

which are line segments in $\mathbb{R}^2$. If $a \in \mathcal{A}$, then $\overline{P}(a)$ is given by

$$\overline{P}(a) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ p(1-p) & (1-p)^2 & p^2 & p(1-p) \\ p(1-p) & p^2 & (1-p)^2 & p(1-p) \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Note that $\overline{P}(a)$ has multiple invariant distributions that are convex combinations of

$$\overline{\sigma}_a^* = (1, 0, 0, 0)$$

and

$$\overline{\sigma}_a^* = (0, 0, 0, 1),$$

which implies that the mean dynamics

$$\begin{bmatrix} \dfrac{da_1}{dt} \\ \dfrac{da_2}{dt} \end{bmatrix} = \begin{bmatrix} \sum_{s \in S} \sigma_a^*(s) u_1(s) - a_1 \\ \sum_{s \in S} \sigma_a^*(s) u_2(s) - a_2 \end{bmatrix} \tag{7.12}$$

for each i, where

$$\sigma_a^* = \sigma_a^* P(a).$$

Recall that the invariant distribution of $P(a)$ must be close to *some* invariant distribution of $\overline{P}(a)$ as $d(h_i, \overline{h}) \to 0$ for $\forall i$. For any invariant distribution of $\overline{P}(a)$,

$$\frac{da_i}{dt} = \sum_{s \in S} \sigma_a^*(s) u_i(s) - a_i > 0 \qquad \forall a_i \leq u_i(D, D) - \rho, \forall \rho > 0.$$

Hence, if the aspiration level of each player i is below $u_i(D, D)$, then the learning dynamics must increase the aspiration level up to $u_i(D, D)$ in finite periods.

It is instructive to see the role of inertia. If there were no inertia so that

$$p = 0,$$

then $\overline{P}(a)$ would have an invariant distribution different from those already identified above:

$$\overline{\sigma}_a^* = \left(0, \frac{1}{2}, \frac{1}{2}, 0\right).$$

This invariant distribution is attained as the outcome alternates between (C, D) and (D, C); in each of these outcomes, the two players synchronously switch to the other action since each player is dissatisfied with the performance of his present choice. By introducing inertia, i.e., assuming $p > 0$, either (C, C) or (D, D) is realized with a positive probability, both of which are absorbing states. As a result, the two players escape almost surely from perpetual coordination failure.

Once the aspiration level approaches $u(D, D)$, the perturbation introduced by the sigmoid function plays an important role. By the definition of the sigmoid function, there exists $\epsilon > 0$ such that

$$\epsilon < h_i(0) < 1 - \epsilon.$$

For an aspiration vector $a = (a_1, a_2)$ in the neighborhood of $u(D, D) = (0, 0)$, the transition matrix $P(a)$ can be approximated by

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ p(1-p) & (1-p)^2 & p^2 & p(1-p) \\ p(1-p) & p^2 & (1-p)^2 & p(1-p) \\ q_1 q_2 & q_1(1-q_2) & q_2(1-q_1) & (1-q_1)(1-q_2) \end{bmatrix}$$

where $q_i = h_i(u_i(s) - a_i) \in (0, 1)$ and $u_i(s) - a_i \simeq 0$. One can easily verify that this transition matrix has a unique invariant distribution

$$(1, 0, 0, 0).$$

As a result, the mean dynamics at aspiration vector $a = (a_1, a_2)$ around the neighborhood of $u(D, D)$ are

$$\frac{da_i}{dt} = u_i(C, C) - a_i > 0 \qquad \forall i = 1, 2.$$

That is, as the aspiration level approaches $u_i(D, D)$, player i experiments more often as (D, D) is realized. As a result, (C, C), which is an absorbing state, is realized with a positive probability.

Our analysis so far indicates that if the initial aspiration level of player i is below $u_i(D, D) = 0$, then the aspiration level should continue to increase so that it can go beyond $u_i(D, D)$ in a finite period of time. Our remaining step is to show that if $a \in A$, then its mean dynamics must converge to the neighborhood of $u(C, C) = (1, 1)$. We essentially repeat the same logic as before. For small $\rho > 0$, $a \in A$ and $a_i > u_i(D,D) + \rho$, the transition matrix $P(a)$ can be approximated by

$$\overline{P}(a) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ p(1-p) & (1-p)^2 & p^2 & p(1-p) \\ p(1-p) & p^2 & (1-p)^2 & p(1-p) \\ p^2 & p(1-p) & p(1-p) & (1-p)^2 \end{bmatrix}$$

which has a unique invariant distribution

$$\overline{\sigma}_a^* = (1,0,0,0).$$

As $P(a)$ converges to $\overline{P}(a)$, so does its invariant distribution. Thus, the mean dynamics can be approximated by

$$\frac{da_i}{dt} = u_i(C,C) - a_i \geq 0,$$

and the equality holds only if $a_i = u_i(C,C)$. This proves that for $\forall a_0$ and $\forall \rho > 0$, $\exists \epsilon > 0$ such that if $d(h_i, \overline{h}) < \epsilon$ for all i, then there exists $T > 0$ such that for all $t \geq T$, $a_i(t) \geq u_i(C,C) - \rho$ where $a(t)$ is induced by (6.9). Then, by invoking Theorem 6.5, we conclude that

$$a_t \to N_\rho(u(C,C))$$

with probability one. Since the only way to achieve an average payoff close to $u_i(C,C)$ is to play (C,C) almost always, this convergence result also implies that $s_t = (C,C)$ for almost all $t \geq 1$.

Albeit tedious, one can apply precisely the same logic to a general coordination game:

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} 1,1 & \ell, g \\ g, \ell & 0,0 \end{pmatrix} \end{array}, \tag{7.13}$$

where $g, \ell < 0$, in order to prove that for all $\rho > 0$, there exists $\epsilon > 0$ such that if $d(h_i, \overline{h}) < \epsilon$ for all $i = 1,2$, then

$$a_t \to N_\rho(u(C,C))$$

with probability one.

7.2 Battle of the sexes

While KMRV, Oechssler (2001) and Posch (2001) also show that the efficient coordination outcome is achieved in the coordination game, their result is restricted to the component games which have a symmetric Pareto efficient (pure strategy) outcome. In contrast, our method can be applied to any 2×2 games that do not have a symmetric Pareto efficient outcome such as the battle of the sexes:

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} 0,0 & 1, 1+g \\ 1+g, 1 & 0,0 \end{pmatrix} \end{array} \tag{7.14}$$

where $g > 0$. Because the analysis of the battle of the sexes follows precisely the same logic as the coordination game, we only sketch the proof. Let V be the set of feasible payoff vectors and partition V into 4 different regions:

$$\begin{aligned} \mathcal{A} &= \{(u_1, u_2) \in V | u_1, u_2 < 1\}, \\ \mathcal{B} &= \{(u_1, u_2) \in V | u_1 < 1, u_2 \geq 1\}, \\ \mathcal{B}' &= \{(u_1, u_2) \in V | u_1 \geq 1, u_2 < 1\}, \\ \mathcal{C} &= \{(u_1, u_2) \in V | u_1, u_2 \geq 1\}, \end{aligned}$$

as depicted in Figure 1. Note that Region $\mathcal{B}'$ is the mirror image of $\mathcal{B}$.

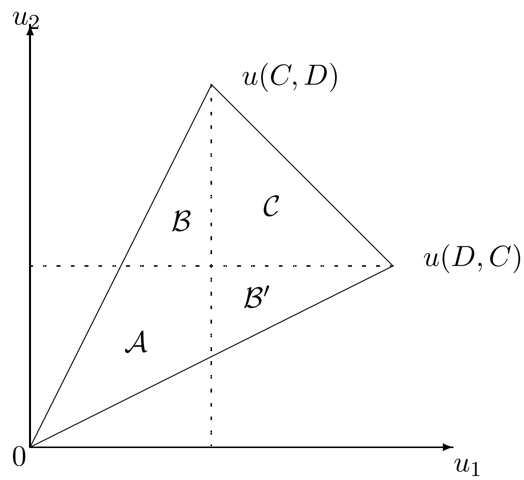

Figure 1: Battle of the Sexes

In Region $\mathcal{A}$, once (C, D) is chosen, the payoff is higher than the aspiration level for both players, and therefore, the two players continue to play (C, D) with a high probability except in the neighborhood of the boundary of $\mathcal{A}$. The same thing can be said of (D, C).

If, on the other hand, either (C, C) or (D, D) is realized, then the aspiration level (weakly) exceeds the payoff for both players. Therefore, with a probability of $2p(1 - p)$, the action pair switches to either (C, D) or (D, C). Thus, if t is sufficiently large, the system moves to one of the off-diagonal action pairs very soon without changing a_t's too much, and stays there for a sufficiently long time.

In Region $\mathcal{B}$ (resp. $\mathcal{B}'$), using the same logic as above, we can verify that, for a sufficiently large t, the system moves to $u(C, D)$ (resp. $u(D, C)$) and stays there with a high probability.

Region $\mathcal{C}$ determines the global stability of the system. The transition matrix in this region is approximated by:

$$\begin{bmatrix} p^2 & p(1-p) & p(1-p) & (1-p)^2 \\ 0 & p & 0 & 1-p \\ 0 & 0 & p & 1-p \\ (1-p)^2 & p(1-p) & p(1-p) & p^2 \end{bmatrix}.$$

Thus, the system moves toward u^* provided that $\lim_{t\to\infty} \gamma_t = 0$ where u^* is calculated from a stationary distribution σ^* of the above transition matrix, i.e.,

$$\sigma^* = \frac{1}{2+4p}(1-p, 2p, 2p, 1+p).$$

Therefore, we have

$$u^* = \sigma^* \cdot u = \frac{p}{1+2p}(2+g, 2+g).$$

Let us perturb the decision rule by replacing the step function $\overline{h}$ by the sigmoid function h_i $(i = 1, 2)$ defined in Section 4. By the definition of h_i, each player is subject to a small probability of making "mistakes". If $u_i^* < 1$, i.e, $pg < 1$, holds, then both $u(C, D)$ and $u(D, C)$ are locally stable. Indeed, consider $u(C, D)$.[12] If a falls in region $\mathcal{B}$, then it is brought back to the neighborhood of $u(C, D)$. Suppose next that a goes away from $u(C, D)$ by a sufficiently small $\varepsilon > 0$ into region $\mathcal{C}$. Then it moves toward u^* by at most $\frac{1+g-pg}{1-pg}\varepsilon$ until it hits the boundary of $\mathcal{B}$ and $\mathcal{C}$, after which a reverts back toward $u(C, D)$ without entering region $\mathcal{C}$. Thus, $u(C, D)$, or to be precise, a stationary point nearby, is locally stable. A symmetric argument can be made of $u(D, C)$.

If, however, $u_i^* > 1$ holds for both $i = 1, 2$, the system shows a different behavioral pattern. The only stable point becomes u^*. Thus, applying Theorem 6.5 to the present game of the battle of the sexes, we obtain the following result where we define

$$\mathcal{K}^{**} = \{u(C, D), u(D, C)\}.$$

Theorem 7.1 *For all $\rho > 0$, there exists $\epsilon > 0$ such that $d(h_i, \overline{h}) < \epsilon$ implies*

$$a_t \to \begin{cases} N_\rho(\mathcal{K}^{**}), & \text{if } pg < 1te \\ N_\rho(u^*), & \text{if } pg > 1 \end{cases}$$

as $t \to \infty$ with probability one. In particular, if p is sufficiently small, then a_t converges to $N_\rho(\mathcal{K}^{**})$.

Proof. See Appendix D.

If $u_i^* < 1$ $(i = 1, 2)$, there is another stationary point of the associated ODE, located in the neighborhood of $(u_1(C, D), u_2(D, C)) = (1, 1)$ as h_i tends to $\overline{h}$ for $i = 1, 2$. However, this point is not locally stable because if a fluctuates a little and falls in region, say, $\mathcal{B}$, the dynamics lead a to $u(C, D)$. Therefore, the stochastic process does not converge to the neighborhood of $(1, 1)$ with any positive probability.

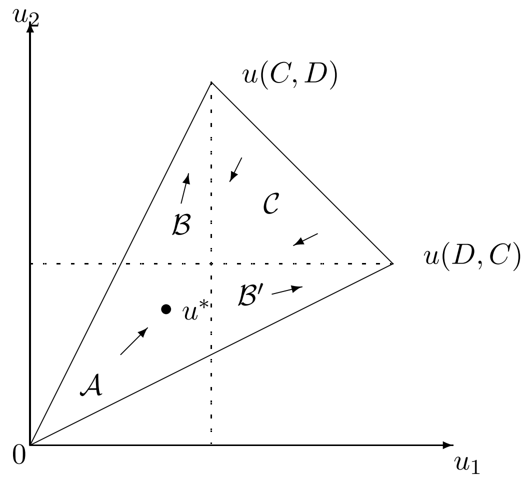

Figure 2: Battle of the Sexes: $u_1^*, u_2^* < 1$

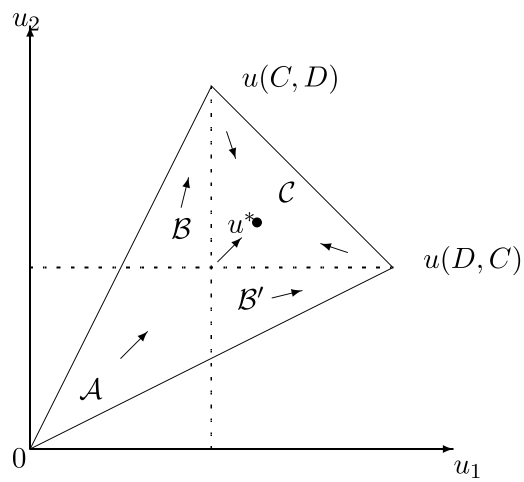

Figure 3: Battle of the Sexes: $u_1^*, u_2^* > 1$

Since we can analyze other 2×2 games by following the same idea, let us simply state the convergence result.

7.3 Game of chicken

Consider the following game of chicken:

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} 1,1 & \ell, 1+g \\ 1+g, \ell & 0,0 \end{pmatrix} \end{array} \tag{7.15}$$

where $g > 0$, $\ell > 0$ and $g + \ell < 1$. We partition the set of feasible payoff vectors into the following regions:

$$\begin{aligned} \mathcal{A} &= \{(u_1, u_2) \in V | \exists s \in S, (u_1, u_2) < u(s)\}, \\ \mathcal{B} &= \{(u_1, u_2) \in V | \ell \leq u_1 < 1, u_2 \geq 1\}. \end{aligned}$$

Like before, Region $\mathcal{B}'$ is the mirror image of $\mathcal{B}$. Note that $\mathcal{A} \cup \mathcal{B} \cup \mathcal{B}' = V$, and the intersection of any pair of these sets is empty.

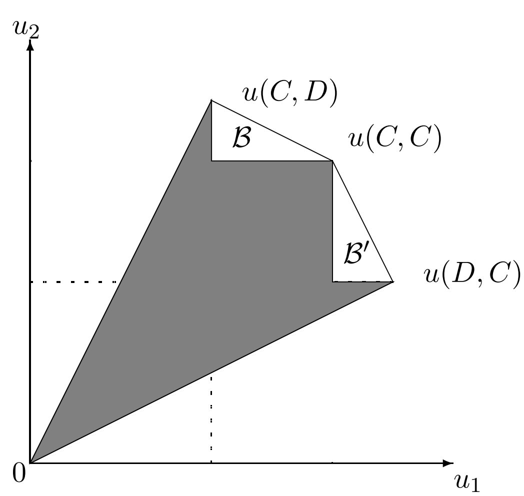

Figure 4: Game of Chicken

Then, we can show that the aspiration vector converges to one of the efficient outcomes.

[12] A stationary point is located slightly away from $u(C, D)$, because h_i is an approximation of the step function. Throughout this paper, we shall "ignore" this difference whenever the meaning is clear from the context.

Theorem 7.2 *For $\forall \rho > 0$, there exists $\overline{\epsilon} > 0$ such that for $\forall \epsilon \in (0, \overline{\epsilon})$, $d(h_i, \overline{h}) < \epsilon$ implies*

$$a_t \to \begin{cases} N_\rho(u(C,C)), & \text{if } 1 + pg > 2\ell \\ N_\rho(\mathcal{K}^*), & \text{if } 1 + pg < 2\ell \end{cases}$$

as $t \to \infty$ with probability one where

$$\mathcal{K}^* = \{u(C,D), u(C,C), u(D,C)\}.$$

If p is sufficiently small, then $\ell = 1/2$ becomes a threshold for the destination of the dynamics. If $\ell < 1/2$ holds, then a_t converges to the neighborhood of $u(C,C)$, while if $\ell > 1/2$, then it converges to one of (the neighborhoods of) $u(C,C)$, $u(C,D)$, and $u(D,C)$.

This conforms to our intuition. If ℓ is small, then the off-diagonal payoffs are "unfair" in the sense that the payoff to one of the players is closer to $u(D,D)$ than to $u(C,C)$. Therefore, the player who is not "well treated" tends to be dissatisfied and moves back and forth between C and D.

7.4 Prisoners' dilemma

Finally, let us examine the prisoners' dilemma game:

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} 1,1 & -\ell, 1+g \\ 1+g, -\ell & 0,0 \end{pmatrix} \end{array} \tag{7.16}$$

where $\ell, g > 0$. We divide V into eight regions:

$$\begin{aligned} \mathcal{A} &= [0,1) \times [0,1), \\ \mathcal{B} &= \{a = (a_1, a_2) \in V | a_1 < 0,\ 0 < a_2 < 1\}, \\ \mathcal{C} &= \{a = (a_1, a_2) \in V | a_1 < 0,\ a_2 \geq 1\}, \\ \mathcal{D} &= \{a = (a_1, a_2) \in V | 0 \leq a_1 < 1,\ a_2 \geq 1\}, \\ \mathcal{E} &= \{a = (a_1, a_2) \in V | a_1 \geq 1,\ a_2 \geq 1\} \end{aligned}$$

and the remaining three regions are the mirror image of $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$ obtained by switching the identity of the players, and are called $\mathcal{B}'$, $\mathcal{C}'$, $\mathcal{D}'$, respectively. Among these eight regions, $\mathcal{E}$ may be empty as depicted in Figure 5 if $u(C,C)$ is located at the Pareto frontier of V.

If the aspiration level pair a is in region $\mathcal{A}$, $\overline{\sigma}_{\mathcal{A}}$ converges to $(1,0,0,0)$ as the sigmoid function h converges to the step function. Therefore, a moves toward $u(C,C)$.

If a is in region $\mathcal{B}$, the unperturbed transition matrix $\overline{P}_{\mathcal{B}}$ is given by

$$\overline{P}_{\mathcal{B}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & p & 0 & 1-p \\ 0 & 0 & p & 1-p \\ 0 & 1-p & 0 & p \end{bmatrix}$$

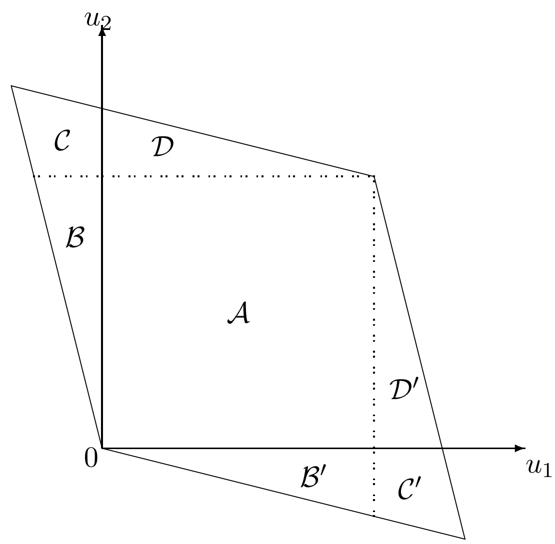

Figure 5: $g < 1,\ \ell < 1$

and any convex combination of $(1,0,0,0)$ and $(0,.5,0,.5)$ is an invariant distribution of $\overline{P}_{\mathcal{B}}$. Thus, $\sigma_a^* \cdot u$ is a convex combination of $u(C,C) = (1,1)$ and $.5u(C,D) + .5u(D,D) = (-\ell/2, (1+g)/2)$. For any invariant distribution, the aspiration level pair will eventually enter region $\mathcal{A}$.[13]

In region $\mathcal{C}$, the unperturbed transition matrix $\overline{P}_{\mathcal{C}}$ is

$$\overline{P}_{\mathcal{C}} = \begin{bmatrix} p & 1-p & 0 & 0 \\ 0 & p & 0 & 1-p \\ 0 & 0 & p & 1-p \\ 0 & 1-p & 1-p & p \end{bmatrix}$$

whose unique invariant distribution is $(0,0,.5,.5)$. Once the aspiration vector enters $\mathcal{C}$, it moves toward $.5u(D,C) + .5u(D,D) = ((1+g)/2, -\ell/2)$. Therefore, the system either reaches the boundary of $\mathcal{C}$ and $\mathcal{D}$ or enters $\mathcal{A}$ in a finite period of time.

Our main interest is whether or not we can sustain the cooperation outcome in the limit. As it turns out, the answer depends upon the size of $g > 0$, which measures the gain from defection against cooperation, or can be interpreted as the temptation for double crossing. It affects the dynamics in quadrant $\mathcal{D}$ (and $\mathcal{D}'$ by symmetry). In $\mathcal{D}$, the transition

[13] Remember that $\overline{P}_{\mathcal{B}}$ is an approximation of $P(a)$ for $\forall a \in \mathcal{B}$, and $P(a)$ has a unique invariant distribution. Because of the multiplicity of invariant distributions of $\overline{P}_{\mathcal{B}}$, the invariant distribution of $P(a)$ is sensitive to the choice of (h_1, h_2). The analysis shows, however, that the asymptotic property of a_t is not sensitive to the choice of (h_1, h_2) since the aspiration level pair enters region $\mathcal{A}$ no matter which direction it might move as discussed in the main text.

matrix $\overline{P}_{\mathcal{D}}$ is given by

$$\overline{P}_{\mathcal{D}} = \begin{bmatrix} p & 1-p & 0 & 0 \\ 0 & p & 0 & 1-p \\ 0 & 0 & p & 1-p \\ (1-p)^2 & p(1-p) & p(1-p) & p^2 \end{bmatrix}.$$

Therefore, the invariant distribution $\overline{\sigma}_{\mathcal{D}}$ is given by

$$\overline{\sigma}_{\mathcal{D}} = \frac{1}{3}(1-p, 1, p, 1).$$

Similarly, we let $\overline{\sigma}_{\mathcal{D}'}$ be given by

$$\overline{\sigma}_{\mathcal{D}'} = \frac{1}{3}(1-p, p, 1, 1).$$

Using this, we define:

$$\begin{aligned} \tilde{u} &= \overline{\sigma}_{\mathcal{D}} u \\ &= \frac{1}{3}\left[(1-p)u(C,C) + u(C,D) + pu(D,C) + u(D,D)\right] \\ &= \left(\frac{1}{3}(1-\ell+pg), \frac{1}{3}(2+g-p-p\ell)\right) \end{aligned}$$

and similarly,

$$\tilde{u}' = \left(\frac{1}{3}(2+g-p-p\ell), \frac{1}{3}(1-\ell+pg)\right).$$

As g increases, both $\tilde{u} = (\tilde{u}_1, \tilde{u}_2)$ and $\tilde{u}' = (\tilde{u}'_1, \tilde{u}'_2)$ are moving away from the main diagonal of V. It is useful to note that

$$\tilde{u}_2 > 1 = u_2(C,C)$$

and

$$\tilde{u}'_1 > 1 = u_1(C,C)$$

if and only if

$$g > 1 + p(1+\ell).$$

For a large ℓ (the loss from being double-crossed), $\tilde{u}$ and $\tilde{u}'$ might fail to be individually rational. In this case, we need to do additional work.

Suppose that $\tilde{u}_1 < 0$. Define

$$\overline{u} = \frac{1}{2}\left(u(D,D) + u(D,C)\right) = \left(\frac{1+g}{2}, -\frac{\ell}{2}\right)$$

as the middle point of $u(D,D)$ and $u(D,C)$. Let $u^* = (0, u_2^*)$ be a point such that the vertical axis

$$\{(u_1, u_2) | u_1 = 0\}$$

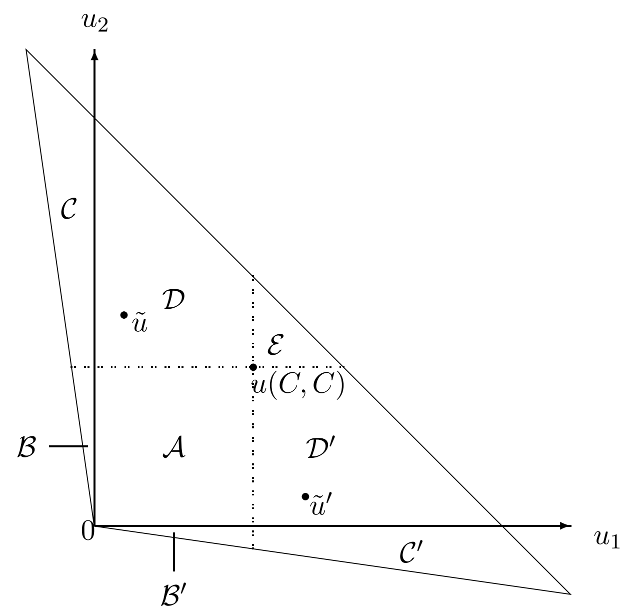

Figure 6: Prisoners' dilemma: $g > 1,\ \ell < 1$

and the line segment connecting $\tilde{u}$ and $\overline{u}$ intersect. Similarly, let $u^{*\prime} = (u_1^{*\prime}, 0)$ be a point at which the horizontal axis and the line segment connecting $\tilde{u}'$ and

$$\overline{u}' = \frac{1}{2}\left(u(D,D) + u(C,D)\right) = \left(-\frac{\ell}{2}, \frac{1+g}{2}\right)$$

intersect. Note that u^* is the mirror image of $u^{*\prime}$ around the main diagonal of V.

Summarizing the above analysis, we obtain the following result.

Theorem 7.3 [1] *Suppose $g \in (0,1)$ and $\ell \in (0,1)$. Then for all $\rho > 0$, there exists $\epsilon > 0$ such that $d(h_i, \mathbf{1}_{x\geq 0}) < \epsilon$ $(i = 1,2)$ implies*

$$a_t \to N_\rho(u(C,C))$$

with probability one.

[2] *Suppose $g > 1$ and $\ell \in (0,1)$. Then for all $\rho > 0$, there exists $\epsilon > 0$ such that $d(h_i, \mathbf{1}_{x\geq 0}) < \epsilon$ $(i = 1,2)$ implies*

$$a_t \to N_\rho(\{\tilde{u}, \tilde{u}'\})$$

with probability one.

[3] *Suppose* $g > 1$ *and* $\ell > 1$.

[3.1] *If* $g^2 + 1 > \ell^2 + \ell$, *then for all* $\rho > 0$, *there exists* $\epsilon > 0$ *such that* $d(h_i, \mathbf{1}_{x \geq 0}) < \epsilon$ $(i = 1, 2)$ *implies*

$$a_t \to N_\rho \left(\{(0, u_2^*), ({u^*}'_1 0)\} \right)$$

with probability one.

[3.2] *If* $g^2 + 1 \leq \ell^2 + \ell$, *then only* $(0, 1)$ *and* $(1, 0)$ *are locally stable points.*

Proof. See Appendix E.

This theorem says that for a sufficiently small p, which is implied by $d(h_i, \mathbf{1}_{x \geq 0}) < \epsilon$, the cooperation outcome can be sustained if the gain from double crossing is modest, i.e., $g < 1$ holds. If one player deviates from the cooperation outcome, the other party, who is double crossed, receives the worst possible payoff in the one shot game. Consequently, the party who is double crossed immediately responds to such a deviation by playing D as well. The resulting outcome is one shot Nash equilibrium, which is Pareto dominated by the cooperation outcome. Dissatisfied with the bad outcome, both players switch to cooperation simultaneously. But, the cooperation cannot last too long. As a result, the limit outcome is a convex combination of (C, C), (D, D) and (D, C) (or (C, D)). Therefore, the initial defector's aspiration level is drawn toward (2+g)/3. Thus, if $g < 1$, such a deviation pushes the defector's aspiration level below the cooperation payoff of one. This force drives the aspiration level toward region $\mathcal{A}$, and therefore, toward $u(C, C)$.

It is straightforward, albeit tedious, to check that $u_2^* < 1$ and ${u^*}'_1 < 1$ if and only if $\ell + \ell^2 > g^2 + 1$. In this case, the loss from being cheated is so large that the convex combination of $u(D, D)$, $u(C, C)$ and $u(D, C)$, for example, might induce the average payoff for player 2 to go below his security level payoff $u(D, D)$. The theorem says that each player can secure his payoff. Consequently, the long run individual payoff of one player becomes precisely his security level payoff. In this case, which is covered by the last part of Appendix E, the analysis of Posch, Pichler, and Sigmund (1999) indicates the possible existence of a stable limit cycle. In all other cases, the set of all locally stable points is the global attractor: for any initial condition, the aspiration vector converges to its neighborhood with probability one in a finite period of time. Although $(0, 1)$ and $(1, 0)$ are locally stable points, we have not been able to check whether or not $\{(0, 1), (1, 0)\}$ are global attractors of the learning dynamics.

8 Aspiration as a weighted average of the two

We have assumed that the aspiration level is the average of one's own payoff as in KMRV. We can apply the same analytic tool to study the models in which the aspiration level is formed in a different manner such as the one that uses the payoff of the other player (Oechssler (2001)) or the average payoff of the players (Posch (2001)). Instead of reproducing Oechssler (2001) and Posch (2001), we shall modify the aspiration formation rule to incorporate the payoff of the other player to illustrate how our analysis can be applied to other models.

Suppose that the aspiration level of each player is calculated as

$$a_{i,t} = a_{i,t-1} + \gamma_t \left[(1-\lambda)u_i(s_t) + \lambda u_j(s_t) - a_{i,t-1}\right] \tag{8.17}$$

where $j \neq i$, and $\lambda \in [0, 1/2]$ is the weight player i puts on the other player's payoff in updating his own aspiration level. Let V^λ be the set of all aspiration vectors in the long run:

$$V^\lambda = \left\{((1-\lambda)a_1 + \lambda a_2, \lambda a_1 + (1-\lambda)a_2) \in \mathbb{R}^2 |\ (a_1, a_2) \in V\right\}.$$

Notice that V^λ shrinks to a segment of the 45 degree line as λ approaches 1/2.

As V^λ changes in λ, the dynamics change accordingly. But, any change in the stochastic process incurred by $\lambda > 0$ is captured by the change of the associated ODE. Thus, we can examine the asymptotic properties of the evolution of aspirations through the associated ODE. First, given an aspiration pair a, use the same transition matrix $P(a)$ as before to calculate the invariant distribution σ_a^* by solving $\sigma_a^* = \sigma_a^* P(a)$. Next, calculate a_i^λ $(i = 1, 2)$ as

$$a_i^\lambda = (1-\lambda)\sigma_a^* u_i + \lambda \sigma_a^* u_j.$$

The mean dynamics at a move the aspiration pair in the direction of $a^\lambda = (a_1^\lambda, a_2^\lambda)$.

We shall examine three repeated games: coordination, the battle of the sexes and the prisoners' dilemma. The analysis of the game of chicken is left for interested readers.

8.1 Coordination games

The analysis is virtually identical to that in Subsection 7.1. The aspiration levels a^λ converge to $u(C, C)$, which is the only stable outcome of the associated ODE.

8.2 Battle of the sexes

See Figure 7. Note that V^λ, which is the solid triangle in Figure 7, is a subset of V, which is the dotted triangle. Also, note that V^λ converges to the 45 degree line through the origin as $\lambda \to 1/2$. Given $\lambda \in (0, 1/2)$, mark each area of V^λ as $\mathcal{A}$, $\mathcal{B}$, $\mathcal{B}'$ and $\mathcal{C}$ as we did in Figure 1.

The mean dynamics over V^λ are fairly easy to analyze. For example, on $\mathcal{C}$, the gradient vector induced by the associated ODE is pointing toward u^*. Thus, if u^* is in $\mathcal{C}$, then u^* is the only stable point of the associated ODE, and therefore, the aspiration vector converges to u^*. Otherwise, we follow precisely the same logic as in Subsection 7.2 in order to show that $\{\tilde{u}^*, \tilde{u}^{**}\}$ is the set of stable points.

8.3 Prisoners' dilemma

By following the same analysis as in Subsection 7.4, we calculate the threshold of g, beyond which cooperation $u(C, C)$ is not achieved. The key payoff profile $\tilde{u}$ is now modified so

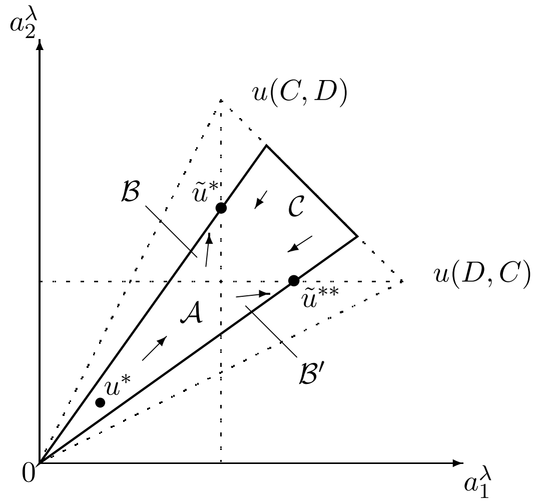

Figure 7: Battle of the Sexes: $\lambda = 1/4$

that we have to pay attention to

$$\tilde{u}^\lambda = (1-\lambda)\tilde{u} + \lambda\tilde{u}'.$$

If this point is in region $\mathcal{D}$ in Figure 5, then $\tilde{u}^\lambda$ is a stable point of the system, and so is its mirror image. If $\tilde{u}^\lambda$ is in region $\mathcal{A}$, then $u(C,C)$ is the unique stable point. In other words, for a sufficiently small p, $u(C,C)$ is the unique stable point if and only if

$$g < 1 + \lambda(1 + g + \ell)$$

holds. Note that the second term on the right hand side is increasing in λ. As λ increases, the threshold for g to sustain cooperation also increases. If $\lambda = 1/2$, the aspiration level is essentially the population average as in Oechssler (2001).

9 Conclusion

We studied the behavior of boundedly rational agents who play an infinitely repeated symmetric 2×2 game according to a simple rule of thumb: each agent continues to play the same action if and only if the action generates one period payoff that exceeds the aspiration level that summarizes the past history according to the average of the past payoff. Using the method of stochastic approximation, we set forth a general approach to this problem and characterized the set of limit points of the learning dynamics for all symmetric 2×2 games.

A couple of remarks are in order. One may wonder if a player who follows this rule is taken advantage of by a rational player. The answer depends on the specific games. In games with multiple individually rational Pareto efficient pure strategy payoffs such as the battle of the sexes, if a rational player knows that the other player is following the above behavior rule, then the rational player can select the equilibrium that favors him. On the other hand, the rational player cannot take advantage of the other in prisoners' dilemma. This is because any defection that leads to the lowest possible payoff for the deceived player prompts him to switch to defection in the next period.

The method of stochastic approximation is applicable, at least in principle, to the class of games with more than two pure strategies. In such a case, we have to specify which action to take when the players switch actions. We may also face a problem of complexity since there will generally be more regions that we have to investigate to characterize the dynamics.

We have not analyzed what would happen if players were randomly matched to play these games. Again, although our approach is applicable in principle to such a situation, we leave it for the future study.

A Proof of Lemma 6.1

Because the complete proof is available in Kushner and Yin (1997) (Theorem 4.2, p.243), we only sketch the proof.

As the first step, choose an arbitrary small $\tau > 0$. Consider (6.7) as $K \to \infty$. By invoking the martingale inequality, one can show that

$$\sum_{t=K+1}^{m(t_K+\tau)-1} \gamma_t \delta M_t \to 0$$

holds for all states except on a null net $\mathcal{N}$. Over $\mathcal{N}^c$, $\{\overline{a}^K(\tau')\}_{K=1}^{\infty}$ is equicontinuous for $\forall \tau' \in [0, \tau]$. By Arzela-Ascoli's lemma, there exists a convergent subsequence. After renumbering the subsequence, let

$$\overline{a}^K(\tau') \to a(\tau') \qquad \forall \tau' \in [0, \tau)$$

uniformly. Since the first term in (6.7) is approximated by the Riemann integration, we have

$$\overline{a}^K(\tau) - \overline{a}^K(0) - \int_0^{\tau} \left[\sigma^*_{a(0)} u - a(\tau')\right] d\tau' - \tau O(\tau) \to 0$$

with probability one. This is true for every convergent subsequence. Since the space is compact, this implies that the sequence itself is convergent, and its limit is $\int_0^{\tau} \left[\sigma^*_{a(0)} u - a(\tau')\right] d\tau' + \tau O(\tau)$. Note that if $\tau > 0$ is sufficiently small, then the trajectory of (7.11) is a good approximation of the sample paths of the stochastic process.

Next, fix an arbitrary $\tau > 0$ and a small $\mu > 0$. Consider a partition of $[0, \tau)$ into $\{[\tau_0, \tau_1), [\tau_1, \tau_2), \ldots, [\tau_{H-1}, \tau_H)\}$ for some H where

$$0 = \tau_0 < \tau_1 < \cdots < \tau_h < \cdots < \tau_H = \tau$$

and

$$\max_{h=1,\ldots,H} \{\tau_h - \tau_{h-1}\} < \mu.$$

Over each time interval $[\tau_{h-1}, \tau_h)$, define

$$a^h(\tau') = a^h(\tau_{h-1}) + \int_0^{\tau' - \tau_{h-1}} \sigma^*_{a^h(\tau_{h-1})} u - a^h(\tau'') d\tau'' \qquad \forall \tau' \in [\tau_{h-1}, \tau_h), \forall h \in \{1, \ldots, H\}$$

with $a^h(0) = \overline{a}^K(0)$ for $\forall K \geq 1$. Let us consider the (deterministic) trajectory $a^h(\tau')$ for $\forall \tau' \in [0, \tau)$, which is essentially the trajectory obtained by "gluing" pieces of the trajectory of the associated ODE over a small time interval. Over the time interval of size μ, the approximation error is bounded by $\mu O(\mu)$. Because the number of intervals in $[0, \tau)$ with length μ increases linearly as μ decreases,

$$\lim_{K \to \infty} \overline{a}^K(\tau') - a^h(\tau') = O(\mu) \qquad \forall \tau' \in [0, \tau).$$

Letting $\mu \to 0$, we obtain the conclusion.

B Proof of Theorem 6.5

In order to prove Theorem 6.5, we verify that our stochastic process satisfies the conditions that appear in Dupuis and Kushner (1989). They are fairly standard, and easy to verify in most application. The numbers in the brackets are the original numbers for the corresponding assumptions in Dupuis and Kushner (1989).[14] We simplify the original version of the theorem by omitting the conditions that are needed only for unbounded problems.

[14] See also Benäim and Weibull (2001).

A1 [2.3]

$$\forall \Delta > 0, \quad \sum_{t=1}^{\infty} e^{-\Delta/\gamma_t} < \infty, \text{ and } \sum_{t=1}^{T} \gamma_t \to \infty \text{ as } T \to \infty;$$

A2 [5.1]

$$\lim_{t_K - t_L \to 0; K, L \to \infty} \frac{\gamma_K}{\gamma_L} = 1.$$

The first condition says that γ_t decreases to 0, but at a proper rate that is roughly slower than $1/t$ but faster than $1/\ln t$. This condition holds for virtually all forms of adaptive learning algorithms with decreasing γ_t which appear in the literature (there is no exception that we know of).

The second condition requires that γ_t should not change drastically. This is a technical condition which is satisfied as long as γ_t changes "smoothly" over time.

Under these conditions, Dupuis and Kushner (1989) proved the following convergence theorem, which is stated in the context of the present model without proof, because it is straightforward to check all the conditions as we have already laid out above.

Theorem B.1 *(Dupuis and Kushner (1989)) Suppose that the stochastic process satisfies [A1] and [A2]. Suppose also that $\mathcal{K}$ is the set of stable solutions of (6.9) with domain of attraction D. If there exists a compact subset D^* in the interior of D such that*

$$\mathcal{K} \subset D^*$$

and

$$a_t \in D^*$$

infinitely many times with probability one, then for all $\rho > 0$,

$$a_t \to N_\rho(\mathcal{K})$$

with probability one.

In our case, the mean dynamics converge to $\mathcal{K}$ starting from any initial condition in $\mathbb{R}^2$. Moreover, for any $\rho > 0$, there exists $T > 0$ such that for $\forall t \geq T$, $a_t \in N_\rho(V)$ where V is the set of all feasible payoff vectors defined in (7.10). Thus, a_t visits the domain of the attraction of the mean dynamics infinitely many times. The conclusion follows.

C Algorithm with constant γ

Although the above theorem is stated with respect to $\{\gamma_t\}$ satisfying [A1], we can apply the same logic to the constant gain algorithm in which a_t evolves according to

$$a_{i,t} = a_{i,t-1} + \gamma \left(u_i(s_t) - a_{i,t-1} \right) \tag{C.18}$$

for some constant $\gamma > 0$ instead of (2.1). In this case, the convergence with probability one does not hold. Instead, we have to use a weaker metric. This version is useful for comparing our model to KMRV.

Theorem C.1 *(Dupuis and Kushner (1989)) Suppose that $\mathcal{K}$ is the set of stable solutions of (6.9) with domain of attraction D. If there exists a compact subset D^* in the interior of D such that*

$$\mathcal{K} \subset D^*$$

and

$$a_t \in D^*$$

infinitely many times with probability one, then for $\forall \rho > 0$ and $\forall T > 0$,

$$\lim_{\gamma \to 0} \Pr \left(\sup_{0 \leq \tau \leq T} d(\overline{a}(\tau), \mathcal{K}) > \rho \right) = 0$$

where $\overline{a}(\tau)$ is the continuous time process obtained by linearly interpolating a_t and $d(\cdot, \cdot)$ is the Hausdorff metric.

D Proof of Theorem 7.1

For the sake of simplicity of expression, we prove the theorem only for the case of symmetric perturbation, i.e., $h_1(x) = h_2(x) = h(x)$ for all x. Also, we parameterize h by ϵ to write h^ϵ though we often suppress the superscript. We assume, again for the sake of simplicity, that:

(i) given $\epsilon > 0$, $h^\epsilon(x) = 1$ if $x \geq \epsilon$ and $h^\epsilon(x) = 0$ if $x \leq -\epsilon$, and that;

(ii) there exists $\delta > 0$ such that given $\epsilon > 0$, $h^\epsilon(x) \in (\delta, 1-\delta)$ holds if $x \in (-\epsilon + \epsilon^2, \epsilon - \epsilon^2)$.

When we let $d(h, \mathbf{1}_{x \geq 0})$ converge to zero, we simply write "ϵ goes to zero." In the following, we presuppose, without rigorous discussion, that ϵ is sufficiently small. In the following, given $\epsilon > 0$ and $a \in \mathbb{R}^2$, we let $N_\epsilon(a) = [a-\epsilon, a+\epsilon] \times [a-\epsilon, a+\epsilon]$.

Lemma D.1 *Suppose $pg < 1$. Then for all $\epsilon > 0$, there exists a symmetric steady state in $N_\epsilon((1,1))$. This steady state is a saddle point.*

Proof. Assume $pg < 1$. In the ϵ-neighborhood of $(1,1)$, we have

$$P(a) \;=\; \begin{bmatrix} p^2 & p(1-p) & p(1-p) & (1-p)^2 \\ 0 & h_1 & 0 & 1-h_1 \\ 0 & 0 & h_2 & 1-h_2 \\ (1-p)^2 & p(1-p) & p(1-p) & p^2 \end{bmatrix}. \tag{D.19}$$

We look for a symmetric equilibrium in this neighborhood. First, solving $\sigma_a = \sigma_a P(a)$, we obtain

$$\sigma_a = \frac{1}{\Sigma}\Big(\frac{(1-h_1)(1-h_2)}{2p}, 1-h_2, 1-h_1, \frac{(1+p)(1-h_1)(1-h_2)}{2p(1-p)}\Big) \tag{D.20}$$

where

$$\Sigma = \frac{(1-h_1)(1-h_2)}{p(1-p)} + 2 - h_1 - h_2,$$

and we abbreviate $h(1-a_i)$ as h_i for $i = 1, 2$. Repeating a similar exercise to the one we do for the prisoners' dilemma, we obtain a symmetric steady state (a^*, a^*) on the line segment connecting $(1-\epsilon, 1-\epsilon)$ and $(1+\epsilon, 1+\epsilon)$.

It must be the case that as ϵ goes to zero, $\sigma_{a^*} \cdot u$ converges to $(1,1)$. Therefore, we have

$$\lim_{\epsilon \to 0} h(1 - a^*(\epsilon)) = 1 - gp(1-p).$$

This implies that $h(1-a^*(\epsilon))$ is bounded away from 0 and 1 in the limit, and therefore, $h'(1-a^*(\epsilon))$ tends to infinity as ϵ goes to zero. Therefore, we obtain

$$\lim_{\epsilon \to 0} \frac{h'(1-a^*(\epsilon))}{\Sigma} \;=\; 1. \tag{D.21}$$

After tedious calculation, we obtain

$$\frac{\partial \sigma_a}{\partial a_1} \cdot u_1 = -\frac{h'(1-a^*)}{\Sigma}\left[\left(1 + \frac{1-h_2}{p(1-p)}\right)\sigma \cdot u_1 - (1+g)\right]$$

and

$$\frac{\partial \sigma_a}{\partial a_1} \cdot u_2 = -\frac{h'(1-a^*)}{\Sigma}\left[\left(1 + \frac{1-h_2}{p(1-p)}\right)\sigma \cdot u_2 - 1\right].$$

$(\partial\sigma_a/\partial a_2) \cdot u_1$ and $(\partial\sigma_a/\partial a_2) \cdot u_2$ are obtained in the same manner. In the limit, as ϵ goes to zero, since $h'(1-a^*(\epsilon))$ tends to infinity, we have

$$\left(\frac{\Sigma}{h'(1-a^*(\epsilon))}\right)^2 (J_{11}J_{22} - J_{12}J_{21}) \to -g^2 < 0, \text{ as } \epsilon \to 0.$$

Thus, a^* is a saddle point.

E Proof of Theorem 7.3

Again we consider a symmetric perturbation only, i.e., $h_1(\cdot) = h_2(\cdot) = h(\cdot)$.

E.1 Lemmata

Lemma E.1 *For all $\epsilon > 0$, $d(h, \mathbf{1}_{x\geq 0}) < \epsilon$ implies that there exists a steady state in $N_\epsilon(u(C,C))$, and that every steady state in $N_\epsilon(u(C,C))$ is locally stable if $g < 1$, and is a saddle point if $g > 1$.*

Proof. Take $\epsilon > 0$ as given. Consider the ϵ-neighborhood of $u(C,C) = (1,1)$. In this neighborhood, we have

$$P(a) \;=\; \begin{bmatrix} h_1 h_2 & h_1(1-h_2) & (1-h_1)h_2 & (1-h_1)(1-h_2) \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \tag{E.22}$$

where we abbreviate $h(1-a_i)$ as h_i $(i = 1,2)$. Solving $\sigma_a = \sigma_a P(a)$, we obtain

$$\sigma_a = \frac{1}{\Sigma}(1, h_1(1-h_2), h_2(1-h_1), 1 - h_1 h_2) \tag{E.23}$$

where

$$\Sigma = 2 + h_1 + h_2 - 3h_1 h_2.$$

This will give us the mean dynamics in this neighborhood:

$$\dot{a} = \Psi(a) = \sigma_a \cdot u - a.$$

$\Psi(a) = 0$ gives us a steady state, which exists in $N_\epsilon(u(C,C))$. We look for a steady state a^* with $a_1^* = a_2^*$. Observe first that due to symmetry, $a_1 = a_2$ implies $\dot{a_1} = \dot{a_2}$. If $a = (1-\epsilon, 1-\epsilon)$, then $h_1 = h_2 = 1$ in (E.23), i.e., $\sigma_a = (1,0,0,0)$. This implies $\dot{a_1} = \dot{a_2} = \epsilon > 0$. On the other hand, consider $a = (1,1)$ if $3 - g + \ell > 0$ and let $a = (1+\epsilon, 1+\epsilon)$. Then $\dot{a_1} = \dot{a_2} < 0$. Since $\sigma_a \cdot u - a$ is continuous in a, there exists $a_1^* = a_2^* \in (1-\epsilon, 1)$ at which $\dot{a_1} = \dot{a_2} = 0$.

Next, for each $\epsilon > 0$, select a steady state $a^*(\epsilon) \in N_\epsilon((1,1))$.[15] Denote $h_i(1-a_i^*(\epsilon))$ by $h_i^*(\epsilon)$ $(i = 1,2)$. In the limit, as ϵ goes to zero, $a^*(\epsilon)$ converges to $(1,1)$. Therefore, it must be the case that $h_i^*(\epsilon)$ $(i = 1,2)$ converges to 1 as well. Since $\sigma_{a^*(\epsilon)} \cdot u_i = a_i^*(\epsilon)$ holds for any $\epsilon > 0$, we have

$$\lim_{\epsilon \to 0} \frac{\left[d\sigma_{a^*(\epsilon)}/d\epsilon\right] \cdot u_i}{da_i^*(\epsilon)/d\epsilon} = 1, \;\; i = 1,2,$$

or

$$-\frac{\partial \sigma_a}{\partial h_1} \cdot u_1 h'(1-a_1^*) - \frac{\partial \sigma_a}{\partial h_2} \cdot u_1 h'(1-a_2^*) \frac{da_2^*}{da_1^*} \;=\; 1, \tag{E.24}$$

$$-\frac{\partial \sigma_a}{\partial h_1} \cdot u_2 h'(1-a_1^*) - \frac{\partial \sigma_a}{\partial h_2} \cdot u_2 h'(1-a_2^*) \frac{da_2^*}{da_1^*} \;=\; 1 \tag{E.25}$$

in the limit as ϵ goes to zero. Due to the assumptions we impose upon $h(\cdot)$, $a_i^*(\epsilon) \in (1-\epsilon, 1-\epsilon+\epsilon^2)$. This implies that $h'(1-a_1^*(\cdot)) = h'(1-a_2^*(\cdot))$ and $da_1^*/da_2^* = 1$ in the limit as ϵ goes to zero. Substituting these equations into (E.24) and (E.25), and solving for $h'(1-a_i^*)$, we obtain

$$[3 - g + \ell] \lim_{\epsilon \to 0} \frac{h'(1-a_i^*(\epsilon))}{\Sigma} \;=\; 1, \;\; i = 1,2 \tag{E.26}$$

where we make use of $\sigma_{a^*} \cdot u = (1,1)$ and $h^* = 1$ in the limit as well.

[15] If there is more than one such equilibrium, we simply choose one of them: $a^*(\epsilon)$ does not have to be the one we found in the last paragraph.

We now examine the stability of a^*. Take the Jacobian of Ψ at a^*:

$$J = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix} = \begin{bmatrix} [\partial\sigma_{a^*}/\partial a_1]\cdot u_1 - 1 & [\partial\sigma_{a^*}/\partial a_2]\cdot u_1 \\ [\partial\sigma_{a^*}/\partial a_1]\cdot u_2 & [\partial\sigma_{a^*}/\partial a_2]\cdot u_2 - 1 \end{bmatrix}. \tag{E.27}$$

Differentiating σ_a with respect to a_1 (or a_2), and evaluating at $a = a^*$, we have

$$\partial\sigma_{a^*}/\partial a_1 = \partial\sigma_{a^*}/\partial a_2 = [(1-3h^*)\sigma_{a^*} - (0, 1-h^*, -h^*, -h^*)]\,\frac{h'(1-a^*)}{\Sigma},$$

and therefore,

$$[\partial\sigma_{a^*}/\partial a_1]\cdot u_1 = [\partial\sigma_{a^*}/\partial a_2]\cdot u_2 = [(1-3h^*)a_i^* + (\ell + (1+g-\ell)h^*)]\,\frac{h'(1-a^*)}{\Sigma}$$

and

$$[\partial\sigma_{a^*}/\partial a_1]\cdot u_2 = [\partial\sigma_{a^*}/\partial a_2]\cdot u_1 = [(1-3h^*)a_i^* - (1+g-(1+g-\ell)h^*)]\,\frac{h'(1-a_i^*)}{\Sigma}$$

where we suppress ϵ.

Note that $h'(1-a_i^*) > 0$. It is immediately verified that $J_{11} + J_{22} < 0$. Therefore, the stability of the system depends on the sign of $J_{11}J_{22} - J_{12}J_{21}$. Since $J_{11} = J_{22} < 0$ and $J_{12} = J_{21} < 0$ hold, its sign is equivalent to that of

$$-J_{11} + J_{12} = -(1+g+\ell)\frac{h'(1-a_i^*(\epsilon))}{\Sigma} + 1 + o(\epsilon).$$

Substituting (E.26) into the above expression, we obtain

$$2\frac{h'(1-a_i^*(\epsilon))}{\Sigma}(1-g) + o(\epsilon)$$

where

$$\lim_{\epsilon\to 0}\frac{o(\epsilon)}{\epsilon} = 0.$$

Since $h'(1-a_i^*(\epsilon))/\Sigma$ is positive, the sign of $J_{11}J_{22} - J_{12}J_{21}$ coincides with that of $1-g$ in the limit as ϵ goes to zero. Hence, a^* is a locally stable state if $g < 1$, and a saddle point if $g > 1$. *Q.E.D.*

The next lemma shows that there is no steady state in the strip around the boundary of $\mathcal{A}$ and $\mathcal{D}$ except around $(1,1)$ and possibly $(0,1)$.

Lemma E.2 **(i)** *Suppose $g < 1$. For all $\rho > 0$, there exists $\epsilon > 0$ such that no steady state exists in $[\epsilon, 1-\rho] \times [1-\epsilon, 1+\epsilon]$.*

(ii) *Suppose $g > 1$. For a sufficiently small $\epsilon > 0$, no steady state exists in $[\epsilon, 1-\epsilon] \times [1-\epsilon, 1+\epsilon]$.*

Proof. In this region, we have

$$P(a) = \begin{bmatrix} h_2 & 1-h_2 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \tag{E.28}$$

where we abbreviate $h(1-a_i)$ as h_i $(i = 1, 2)$. Solving $\sigma_a = \sigma_a P(a)$, we obtain

$$\sigma_a = \frac{1}{\Sigma}(1, 1-h_2, 0, 1-h_2) \tag{E.29}$$

where

$$\Sigma = 1 + 2(1-h_2).$$

Therefore, we have

$$\sigma_a \cdot u_1 = \frac{1-\ell(1-h_2)}{1+2(1-h_2)}, \tag{E.30}$$

$$\sigma_a \cdot u_2 = \frac{1+(1+g)(1-h_2)}{1+2(1-h_2)}. \tag{E.31}$$

In a steady state, it must be the case that

$$1-\epsilon \le \frac{1+(1+g)(1-h_2)}{1+2(1-h_2)} \le 1+\epsilon$$

holds. If $g \neq 1$ holds, then h_2 has to go to 1 as ϵ tends to zero. This implies that

$$\sigma_a \cdot u_1 = \frac{1-\ell(1-h_2)}{1+2(1-h_2)} \to 1$$

as $\epsilon \to 0$. This proves the result for the case of $g < 1$.

Also, our assumption that $h(\cdot)$ and h_2 are close to one implies that

$$\sigma_a \cdot u_2 = a_2 = 1-\epsilon + o(\epsilon) < 1$$

in a steady state. However, if $g > 1$, then $\sigma_a \cdot u_2 > 1$ holds by (E.31). This is a contradiction. Thus, there is no steady state in $[\epsilon, 1-\epsilon] \times [1-\epsilon, 1+\epsilon]$. *Q.E.D.*

Lemma E.3 *Suppose $g < 1$. For all $\rho > 0$, there exists $\epsilon > 0$ such that no steady state exists outside $N_\rho((1,1))$.*

Proof. In the ϵ-boundary between $\mathcal{C}$ and $\mathcal{D}$, $[-\epsilon, \epsilon] \times [1-\epsilon, \infty) \cap V$, we have

$$P(a) = \begin{bmatrix} h_2 & 1-h_2 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1-h_1 & 0 & h_1 & 0 \end{bmatrix}. \tag{E.32}$$

Solving $\sigma_a = \sigma_a P(a)$, we obtain

$$\sigma_a = \frac{1}{\Sigma}(1-h_1, (1-h_1)(1-h_2), h_1(1-h_2), 1-h_2) \tag{E.33}$$

where

$$\Sigma = 3 - h_1 - 2h_2. \tag{E.34}$$

Thus, $g < 1$ implies

$$\sigma_a \cdot u_2 = \frac{1}{\Sigma}\left[1-h_1+(1+g)(1-h_1)(1-h_2) - \ell h_1(1-h_2)\right] < 1. \tag{E.35}$$

Hence, $\Psi(\cdot) < 0$ holds for a sufficiently small $\epsilon > 0$. *Q.E.D.*

Lemma E.4 *Suppose $g < 1$. For all $\rho > 0$, there exists $\epsilon > 0$ such that no closed orbit goes out of $N_\rho((1,1))$.*

Proof. Consider the region with $a_2 > a_1$. The opposite case is treated in a symmetric manner. Take an arbitrary $\rho > 0$ that is not too large. By Lemma E.1, we can find $\epsilon > 0$ such that every steady state is contained in $N_{\rho/3}((1,1))$. Take $\epsilon > 0$ sufficiently small so that

$$M\epsilon < \rho/3$$

holds where $M > \frac{2+\ell}{1-g}$.

From Poincaré's index theory, if there exists a limit cycle, it must contain at least one steady state other than saddle points. Therefore, every limit cycle passes through $N_{\rho/2}((1,1))$. In $\mathcal{A}$ outside the ϵ-boundary, the process moves upward (toward $(1,1)$), while in $\mathcal{D}$ outside the ϵ-boundary, the process moves downward (toward $\tilde{u}$). Therefore, if the cycle leaves $N_{\rho/2}((1,1))$, it must pass through the ϵ-strip between the two regions.

Take $a = (a_1, a_2)$ with $a_1 \in [1-\rho, 1-\rho/3]$ and $a_2 \in [1-\epsilon, 1+\epsilon]$. Then we have

$$\Psi_2(a) \leq \epsilon - \frac{(1-g)(1-h_2)}{1+2(1-h_2)},$$

and

$$\Psi_1(a) \geq M\epsilon - \frac{(2+\ell)(1-h_2)}{1+2(1-h_2)}.$$

Therefore, by the choice of M, $\Psi_1(a) > 0$ whenever $\Psi_2(a) \geq 0$. If both $\Psi_1(a)$ and $\Psi_2(a)$ are negative, then

$$\frac{\Psi_2(a)}{\Psi_1(a)} \geq \frac{1}{M}.$$

Thus, a_1 can be away from 1 by at most $2M\epsilon + \rho/3 < \rho$. Hence, the orbit never goes out of $N_\rho((1,1))$.

Q.E.D.

E.2 $g > 1$ and $\ell > 1$

Lemma E.5 *Suppose $g > 1$ and $\ell > 1$.*

(i) *Suppose $\ell^2 - g^2 < -\ell + 1$. Then, in the limit as ϵ goes to zero, the only steady states are $(0, u_2^*)$ and $(u_1^*, 0)$.*

(ii) *Suppose $\ell^2 - g^2 > -\ell + 1$. Then, in the limit as ϵ goes to zero, the only steady states are $(0,1)$ and $(1,0)$.*

All of these steady states in (i) and (ii) are locally stable.

Proof. Let us prove first that if there is a stationary point in this case, it must be locally stable. The existence of a stationary point in this case is a little involved, and will be proved shortly after we establish the stability property.

From the proof of Lemma E.3, the transition matrix in the ϵ-boundary between $\mathcal{C}$ and $\mathcal{D}$, $[-\epsilon, \epsilon] \times [1-\epsilon, \infty) \cap V$ is given by (E.32). From (E.33) with (E.34), we obtain

$$\sigma_a \cdot u_1 = \frac{1}{\Sigma}\left[1 - h_1 - \ell(1-h_1)(1-h_2) + (1+g)h_1(1-h_2)\right] \tag{E.36}$$

and

$$\sigma_a \cdot u_2 = \frac{1}{\Sigma}\left[1 - h_1 + (1+g)(1-h_1)(1-h_2) - \ell h_1(1-h_2)\right]. \tag{E.37}$$

If $(0,1)$ is a steady state in the limit, then it must be the case that

$$h_1 = \frac{g-1}{1+g+\ell}$$

and

$$h_2 = 1 - \frac{2+\ell}{(\ell - g + 1)(1+g+\ell)}.$$

Therefore, $(0,1)$ is a steady state in the limit of $\epsilon \to 0$ if $\ell^2 - g^2 > -\ell + 1$ holds, and it is not if $\ell^2 - g^2 < -\ell + 1$ holds. We divide the proof into two cases.

Case (i) $\ell^2 - g^2 < -\ell + 1$: In the region $[-\epsilon, \epsilon] \times [1+\epsilon, \infty) \cap V$, which is the ϵ-boundary except around $(0,1)$, we have $h_2 = 0$ in (E.32). Thus, $\sigma_a \cdot u_1 = 0$ in the limit implies

$$h_1 = \frac{g-1}{g+\ell}$$

at the steady state. Using this, we obtain

$$u_2^* = \sigma_a \cdot u_2 = 1 + \frac{g^2 - \ell^2 - \ell + 1}{3g + 2\ell + 1},$$

which is greater than 1 due to the condition $\ell^2 - g^2 < -\ell + 1$.

Note that the condition $\ell^2 - g^2 < -\ell + 1$ implies that $h(\cdot)$ at the steady state is bounded away from 0 and 1 in the limit, and therefore, $h'(\cdot)$ tends to infinity as ϵ goes to zero. Therefore, we obtain

$$\begin{aligned} J_{11}\frac{\Sigma}{h'} &\to -(\ell-1) < 0, \text{ as } \epsilon \to 0, \\ J_{22} &= -1 < 0, \\ J_{12} &= 0, \\ J_{21}\frac{\Sigma}{h'} &\to 2 + g + \ell > 0, \text{ as } \epsilon \to 0. \end{aligned}$$

Thus, $J_{11} + J_{22} < 0$ and $J_{11}J_{22} - J_{12}J_{21} > 0$, and therefore, $(0, u_2^*)$ is stable. By a symmetric argument, $(u_1^*, 0)$ is stable, too.
Case (ii) $\ell^2 - g^2 > -\ell + 1$: Note that the condition $\ell^2 - g^2 > -\ell + 1$ implies that $h(\cdot)$ at the steady state is bounded away from 0 and 1 in the limit, and therefore, $h'(\cdot)$ tends to infinity as ϵ goes to zero. Therefore, we obtain

$$\begin{aligned} J_{11}\frac{\Sigma}{h'} &\to -\frac{1+g}{\ell - g + 1} < 0, \\ J_{22}\frac{\Sigma}{h'} &\to 0, \\ J_{12}\frac{\Sigma}{h'} &\to -(\ell - g + 1) < 0, \\ J_{21}\frac{\Sigma}{h'} &\to \frac{2+\ell}{\ell - g + 1} > 0, \end{aligned}$$

as ϵ goes to zero. Thus, $J_{11} + J_{22} < 0$ and $J_{11}J_{22} - J_{12}J_{21} > 0$, and therefore, $(0, 1)$ is stable. By a symmetric argument, $(1, 0)$ is stable, too.

E.3 Proof of the theorem

If $g < 1$, then the ODE vanishes at a single point in the neighborhood of $(1, 1)$, which we demonstrated is locally stable. The phase diagram of the ODE reveals that starting from any initial condition, the trajectory of the ODE must converge to the stationary point. That is, the locally stable point of the ODE is the globally stable point. By applying Theorem 6.5, we conclude that if $g < 1$, then a_t converges to the small neighborhood of $(1, 1)$ with probability one. This proves the first part of the theorem.

If $g > 1$, then the stationary solution of the ODE in the neighborhood of $(1, 1)$ is a saddle point. We have to show that the aspiration vector converges to either one of the locally stable points in the limit, which are $\{\tilde{u}, \tilde{u}'\}$ if $g > 1$ and $\ell < 1$, and $\mathcal{K}^*$ if $g > 1$ and $\ell > 1$. From the previous analysis, we know that the ODE vanishes precisely at three points: two locally stable points and one saddle point in the neighborhood of $(1, 1)$.

To simplify the notation, let us regard $(1, 1)$ as the saddle point, and let $\mathcal{S}^* = \{s_1^*, s_2^*, (1, 1)\}$ be the three points where the ODE vanishes. If $g > 1$ and $\ell > 1$, then $s_1^* = (0, \max[u_2^*, 1])$, and $s_2^* = (\max\left[{u^*}'_1, 1\right], 0)$, and if $g > 1$ and $\ell < 1$, then $s_1^* = \tilde{u}$ and $s_2^* = \tilde{u}'$.

Suppose that $g > 1$ and $\ell < 1$. Then, $s_1^* = \tilde{u}$ and $s_2^* = \tilde{u}'$. It is straightforward to verify that starting from any initial condition, the trajectory of the ODE must converge to the neighborhood of $\{s_1^*, s_2^*\}$. Then, Theorem 6.5 implies that $a_t \to \{s_1^*, s_2^*\}$ with probability one. If $g > 1$ and $\ell > 1$, but $s_1^* = (0, u_2^*)$ and $s_2^* = ({u^*}'_1, 0)$, then we can invoke exactly the same logic to prove that $a_t \to \{s_1^*, s_2^*\}$ with probability one.

We now prove the theorem for the case where $g > 1$ and $\ell > 1$ so that $s_1^* = (0, 1)$ and $s_2^* = (1, 0)$:

$$\ell^2 - g^2 > -\ell + 1.$$

We shall focus on the stability property of s_1^*, because the other case follows from the symmetric argument.

Note that s_1^* is the corner of $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ and $\mathcal{D}$. We can write the transition matrix as

$$\begin{bmatrix} 1-h_2 & h_2 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ h_1 & 0 & 1-h_1 & 0 \end{bmatrix}$$

where h_i is continuous $(i = 1, 2)$; $h_1(x) = 0$ if $x < -\epsilon$, $h_1(x) = 1$ if $x > \epsilon$, and h_1 is strictly increasing over $[-\epsilon, \epsilon]$; and $h_2(x) = 0$ if $x < 1-\epsilon$, $h_2(x) = 1$ if $x > 1+\epsilon$, and h_2 is strictly increasing over $[1-\epsilon, 1+\epsilon]$. For example, in the area of $\mathcal{D}$ which is away from its boundary by more than ϵ, the transition matrix is

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

because $h_1 = h_2 = 1$. If $h_1 > 0$ or $h_2 > 0$, then the invariant distribution is

$$\frac{1}{h_1 + 2h_2} \left(h_1, h_1 h_2, h_2(1-h_1), h_2 \right).$$

If $h_1 = 0$ and $h_2 = 0$ (which is the case in the "interior" of $\mathcal{B}$), the invariant distributions are

$$\left\{ \left(\lambda, 0, \frac{1-\lambda}{2}, \frac{1-\lambda}{2} \right) \mid \lambda \in [0,1] \right\}.$$

Note that this case does not arise if we assume that the value of the sigmoid function is contained in $(0, 1)$. But at the same time, the existence of multiple invariant distributions reveals that the sequence of the unique invariant distributions induced by the sigmoid functions might be very sensitive to the choice of the sigmoid function. Thus, we have to pay special attention to the case where $h_1 = h_2 = 0$, or equivalently, $\{(a_1, a_2) | a_1 \le -\epsilon, a_2 \le 1-\epsilon\}$.

If $h_1 + 2h_2 > 0$, then the ODE is

$$\dot{a}_1 = \frac{1}{h_1 + 2h_2} \left(h_1 - \ell h_1 h_2 + (1+g) h_2 (1-h_2) \right) - a_1 \tag{E.38}$$

$$\dot{a}_2 = \frac{1}{h_1 + 2h_2} \left(h_1 + (1+g) h_1 h_2 - \ell h_2 (1-h_2) \right) - a_2. \tag{E.39}$$

Clearly, at the stationary point s_1^*, $\dot{a}_1 = \dot{a}_2 = 0$.

Our task is to extract information about

$$\dot{A}_1 = \{(a_1, a_2) | \dot{a}_1 = 0\}$$

and

$$\dot{A}_2 = \{(a_1, a_2) | \dot{a}_2 = 0\}.$$

Let us examine $\dot{A}_1$. Since h_2 is an increasing function of a_2,

$$\frac{\partial \dot{a}_1}{\partial a_2} \le 0 \tag{E.40}$$

and if $h_2' > 0$, then the strict inequality holds. Thus, if $\dot{a}_1 = 0$ at (a_1, a_2), then for any $(a_1, a_2') \notin \dot{A}_1$, $\dot{a}_1 < 0$ if $a_2' > a_2$, and vice versa.

Consider the area where $h_2 = 1$, which is the case if $a_2 \ge 1+\epsilon$. A simple calculation shows that

$$\dot{a}_1 = \frac{1 + g - (\ell + g) h_1}{h_1 + 2} - a_1 = 0$$

at $\dot{A}_1$. Since h_1 is an increasing function of a_1, this equation has a unique solution that satisfies the equality, which implies that $\dot{A}_1$ is a vertical line if $h_2 = 1$:

$$(a_1, a_2), (a_1', a_2') \in \dot{A}_1, a_1 = a_1' \quad \text{if} \quad a_2, a_2' \ge 1+\epsilon. \tag{E.41}$$

Similarly, let us consider the area where $h_1 = 1$, which covers the area where $a_1 \geq \epsilon$. If $h_1 = 1$, then

$$\dot{a}_1 = \frac{1 - \ell h_2}{1 + 2h_2} - a_1,$$

which is a strictly decreasing function of h_2:

$$\frac{\partial \dot{a}_1}{\partial a_2} < 0. \tag{E.42}$$

In particular, if $a_1 = \epsilon$, then $h_2 = (1-\epsilon)/(\ell + 2\epsilon)$, and if $a_1 = 1$, then $h_2 = 0$. Thus, $\dot{A}_1$ is located in a narrow band of $[\epsilon, 1] \times [1-\epsilon, 1+\epsilon]$. Combining (E.40), (E.41) and (E.42), we conclude that $\dot{A}_1$ is contained in a narrow band of

$$[-\epsilon, \epsilon] \times [1-\epsilon, \infty) \cup [-\epsilon, \infty) \times [1-\epsilon, 1+\epsilon].$$

Moreover, $\dot{A}_1$ is "decreasing": if $\dot{A}_1$ is not a vertical line, then it has a negative slope.

Next, we examine $\dot{A}_2$. It is easy to verify that

$$\frac{\partial \dot{a}_2}{\partial a_1} > 0. \tag{E.43}$$

If $h_1 = 0$ and $h_2 > 0$, or equivalently, $a_1 < -\epsilon$ and $a_2 > 1 + \epsilon$, then

$$\dot{a}_2 = -\frac{\ell}{2} - a_2 < 0. \tag{E.44}$$

If $h_1 > 0$ and $h_2 = 0$, then $a_1 > 1 - \epsilon$, $a_2 \leq 1 - \epsilon$ and

$$\dot{a}_2 = 1 - a_2 \geq \epsilon > 0. \tag{E.45}$$

If $h_1 = 1$ and $1 - \epsilon < h_2 \leq 1 + \epsilon$, then $a_1 \geq \epsilon$, $1 - \epsilon < a_2 \leq 1 + \epsilon$ and

$$\dot{a}_2 = \frac{(1+g)h_2 + 1}{1 + 2h_2} - a_2.$$

It would be helpful to understand the structure of the sigmoid function h_2 over $[1-\epsilon, 1+\epsilon]$. Because h_2 has to change from 0 to 1 over a small interval, its derivative must be very large over almost all parts of $[1-\epsilon, 1+\epsilon]$. In particular, $h'(0) \uparrow \infty$. Thus, if $\partial \dot{a}_2/\partial a_2 < 0$ at $a_2 = 1-\epsilon$, then the derivative must become positive before a_2 becomes equal to 0. Thus, the minimum of $\dot{a}_2$ over $[1-\epsilon, 1+\epsilon]$ must be located at $a_2 \in (1-\epsilon, 0)$. Since $\partial \dot{a}_2/\partial a_2 \geq -1$, the minimum of $\dot{a}_2$ cannot be smaller than

$$\frac{1 + (1+g)h_2(1-\epsilon)}{1 + 2h_2(1-\epsilon)} - (1-\epsilon) = \epsilon.$$

Thus, for a given $g > 1$, we can choose $\epsilon > 0$ sufficiently small so that

$$\frac{1 + (1+g)h_2(1+\epsilon)}{1 + 2h_2(1+\epsilon)} - (1+\epsilon) = \frac{2+g}{3} - (1+\epsilon) > 0.$$

Then, it follows that

$$\dot{a}_2 > 0$$

if $a_1 \geq \epsilon$ and $a_2 \in [1-\epsilon, 1+\epsilon]$.

So far, we have shown that $\dot{a}_2 < 0$ over

$$(-\infty, -\epsilon] \times (1-\epsilon, \infty) \tag{E.46}$$

and that $\dot{a}_2 > 0$ over

$$([-\epsilon, \infty) \times (-\infty, 1-\epsilon]) \cup ([1+\epsilon, \infty) \times [1-\epsilon, 1+\epsilon]). \tag{E.47}$$

Since the gradient vector is a continuous function of the aspiration vector, $\dot{A}_2$ must be located outside of (E.46) and (E.47). Combining this observation with (E.40), (E.41) and (E.42), we conclude that *both* $\dot{A}_1$ and $\dot{A}_2$ can exist only in

$$\{(a_1, a_2) \mid -\epsilon < a_1 < \epsilon, a_2 > 1 - \epsilon\}. \tag{E.48}$$

Thus, if there is a locally stable point, then it must be located in (E.48).

For convenience, let us regard Figure 5 as a "map." Note that $\dot{A}_2$ moves to the southwest region from (E.48), while $\dot{A}_1$ moves to the southeast region from (E.48). Thus, to prove that there exists a locally stable point, it suffices to show that over (E.48), $\dot{A}_2$ is located to the east side of $\dot{A}_1$.

We know that in $\dot{A}_1$, if $a_2 \geq 1+\epsilon$, then $h_2 = 1$ and therefore,

$$\frac{h_1 - \ell h_1 + (1+g)(1-h_2)}{h_1+2} - a_1 = 0.$$

Hence,

$$h_1 = \frac{1+g-2a_1}{\ell+g-a_1}.$$

Substituting h_1 and setting $h_2 = 1$ into $\dot{a}_2$ evaluated at $(a_1, a_2) \in \dot{A}_1$, we have

$$\dot{a}_2 = \frac{1}{h_1+2}\left[\frac{(1+g-2a_1)(2+\ell+g-a_2)}{\ell+g-a_1} - \ell - 2a_2\right].$$

Our task is to evaluate the sign of the term inside the brackets. Recall that we are investigating the gradient vector along $\dot{A}_1$ in the area of $a_1 \in [-\epsilon, \epsilon]$ and $a_2 \geq 1+\epsilon$. Thus, $a_1 \leq \epsilon$ and $a_2 \geq 1+\epsilon$. Hence,

$$\frac{(1+g-2a_1)(2+\ell+g-a_2)}{\ell+g-a_1} - \ell - 2a_2 \leq \frac{(1+g-2\epsilon)(\ell+g+1-\epsilon)}{\ell+g-\epsilon} - \ell - 2 - 2\epsilon.$$

Note that if $\epsilon = 0$, the right hand side of the inequality becomes

$$\frac{(1+g)(\ell+g+1)}{\ell+g} - \ell - 2,$$

which is negative if and only if

$$g^2 - \ell^2 - \ell + 1 < 0,$$

which is precisely the case we are investigating. Thus, for a sufficiently small $\epsilon > 0$,

$$\dot{a}_2 < 0$$

for $(a_1, a_2) \in \dot{A}_2 \cap \{a_2 \geq 1+\epsilon\}$. By (E.43), over the area located in the left hand side of $\dot{A}_2$, $\dot{a}_2 < 0$. Thus, $\dot{A}_1$ is located in the left hand side of $\dot{A}_2$ if $a_1 \geq 1+\epsilon$.

Notice that a part of $\dot{A}_2$ is located in the area where $\dot{a}_1 < 0$, while the rest is located in the area where $\dot{a}_1 > 0$ (i.e, $a_1 \leq 1-\epsilon$). Since the gradient changes continuously with respect to the aspiration vector,

$$\dot{A}_1 \cap \dot{A}_2 \neq \emptyset,$$

which proves that a stationary point exists.

References

Abreu, D., and A. Rubinstein (1988): "The Structure of Nash Equilibrium in Repeated Games with Finite Automata," *Econometrica*, 56(6), 1259–1281.

Aumann, R., and S. Sorin (1989): "Bounded Rationality and Cooperation," *Games and Economic Behavior*, 1, 5–39.

Aumann, R. J. (1959): "Acceptable Points in General Cooperative n-person Games," in *Contributions to Theory of Games IV*, ed. by A. W. Tucker, and R. D. Luce, Annals of Mathematics Studies 40. Princeton University Press.

Benäim, M., and J. Weibull (2001): "Deterministic Approximation of Stochastic Evolution in Games," *forthcoming in Econometrica*.

Carmichael, H. L., and W. B. McLeod (1997): "Gift Giving and the Evolution of Cooperation," *International Economic Review*, 38(3), 485–509.

Dupuis, P., and H. J. Kushner (1989): "Stochastic Approximation and Large Deviations: Upper Bounds and w.p.1 Convergence," *SIAM Journal of Control and Optimization*, 27, 1108–1135.

Evans, G. W., and S. Honkapohja (2001): "Expectations and the Stability Problem for Optimal Monetary Policies," University of Oregon.

Gilboa, I., and D. Schmeidler (1995): "Case-Based Decision Theory," *Quarterly Journal of Economics*, 15, 1–26.

——— (2000): "Reaction to Price Changes and Aspiration Level Adjustments," Tel Aviv University.

——— (2001): *A Theory of Case-Based Decisions*. Cambridge University Press.

Karandikar, R., D. Mookherjee, D. Ray, and F. Vega-Redondo (1998): "Evolving Aspirations and Cooperation," *Journal of Economic Theory*, 80(2), 292–331.

Kushner, H. J., and D. S. Clark (1978): *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. Springer-Verlag.

Kushner, H. J., and G. G. Yin (1997): *Stochastic Approximation Algorithms and Applications*. Springer-Verlag.

Ljung, L. (1977): "Analysis of Recursive Stochastic Algorithms," *IEEE Transactions on Automatic Control*, pp. 551–575.

Marcet, A., and T. J. Sargent (1989): "Convergence of Least Squares Learning Mechanisms in Self Referential Linear Stochastic Models," *Journal of Economic Theory*, 48, 337–368.

Matsui, A. (2000): "Expected utility and case-based reasoning," *Mathematical Social Sciences*, 39, 1–12.

Napel, S. (2003): "Aspiration Adaptation in the Ultimatum Minigame," *forthcoming in Games and Economic Behavior*.

Oechssler, J. (2001): "Cooperation as a Result of Learning with Aspiration," forthcoming in Journal of Economic Behavior and Organization.

Posch, M. (2001): "Win Stay, Lose Shift or Imitation — Only the Choice of the Peers Counts," University of Vienna.

Posch, M., A. Pichler, and K. Sigmund (1999): "The Efficiency of Adapting Aspiration Levels," in *Proceedings of Royal Society*, vol. 266, pp. 1427–1435, London.

Rubinstein, A. (1986): "Finite Automata Play Repeated Prisoners Dilemma," *Journal of Economic Theory*, 39(1), 83–96.

Sandholm, W., and J. Hofbauer (2002): "On the Global Convergence of Stochastic Fictitious Plays," *Econometrica*, 70, 2265–2294.

Sarin, R., and F. Vahid (1999): "Payoff Assessment without Probabilities: A Simple Dynamic Model of Choice," *Games and Economic Behavior*, 28(2), 294–309.

Simon, H. (1987): *Models of Man*, Continuity in Administrative Science. Ancestral Books in the Management of Organizations. Garland Publisher.

Thorndike, E. L. (1911): *Animal Intelligence: Experimental Studies*. MacMillan, New York.

Vega-Redondo, F. (1997): "Evolution of Walrasian Behavior," *Econometrica*, 65(2), 375–384.

Winter, S. G. (1971): "Satisficing, Selection and the Innovating Remnant," *Quarterly Journal of Economics*, 85(2), 237–261.

Woodford, M. D. (1990): "Learning to Believe in Sunspots," *Econometrica*, 58, 277–307.