Unique Stationary Behavior

Heller, Yuval and Mohlin, Erik

University of Oxford

13 August 2015

Unique Stationary Behavior

Yuval Heller and Erik Mohlin*
Department of Economics, University of Oxford.

August 14, 2015

Abstract

We study environments in which agents from a large population are randomly matched to play a one-shot game, and, before the interaction begins, each agent observes noisy information about the partner's aggregate behavior. Agents follow stationary strategies that depend on the observed signal. We show that every strategy distribution admits a unique behavior if each player observe on average less than action of his partner. On the other hand, if each player observes on average more than one action, we show that there exists a stationary strategy that admits multiple consistent outcomes.

Keywords: Markovian process, Random matching. **JEL Classification:** C72, C73, D83.

1 Introduction

We consider an infinite population of agents who are randomly matched into pairs to play a symmetric one-shot game. Before playing the game, each agent privately makes a finite number of independent *observations* sampled from his partner's aggregate behavior. Each such observation includes the realized pure action played by the partner, and may also include the action played by his partner's opponent (i.e., the observation may be an action profile). We assume that each agent follows a stationary (Markovian) *strategy*: a mapping that assigns a mixed action to each possible observation, and which does not depend on the player's own private history or on time. We interpret a (finite support) *strategy distribution* as describing a population in which different groups in the population follow different strategies.[1]

For concreteness consider the following two examples:

1. Each agent uses the signals about the partner's past behavior, to infer the partner's likely behavior in the current interaction, and best-replies to this belief about the partner's behavior.

2. Agents interact in the Prisoner's Dilemma game, and some agents follow a strategy of indirect reciprocity: they cooperate if the partner is observed to cooperate with a sufficiently high frequency. The study of these strategies was pioneered in Nowak & Sigmund (1998) (where it is called "image scoring"), and recently it has extensively analyzed by us in a separate paper (Heller & Mohlin, 2015).

*Email: yuval.heller@economics.ox.ac.uk and erik.mohlin@economics.ox.ac.uk.

[1] See Rosenthal (1979) for an early model of agents who follow stationary strategies and are randomly matched to play one-shot games; see Bhaskar *et al.* (2013) for a foundation for the use of stationary strategies (in a somewhat different setup).

An *outcome* is a mapping that describes the mixed action played by each group in the population conditional on being matched with individuals from each other group. An outcome is *consistent* with the strategy distribution, if for any two strategies in the support of the strategy distribution, it is the case that if the observations are sampled from the outcome, then the induced play coincides with the mixed actions described by the outcome.

The first question when analyzing such interactions is whether the strategy distribution uniquely determines the outcome, or whether there might exist multiple consistent outcomes, such that history determines which one is played. In this note we investigate what conditions on the observation structure that implies that every distribution of stationary strategies induces a unique outcome.[2]

Theorem 1 shows that every strategy uniquely determines the outcome if and only if the expected number of actions that each agent observes is less than one. Moreover, when the expected number of observed actions is less than one, then there is a global convergence at an exponential rate to the unique consistent outcome from any initial outcome.

The intuition can be grasped by considering an environment in which each agent observes a single action with probability p, and observes nothing otherwise. In the former case, the agents plays the observed action, and in the latter case he plays an arbitrary mixed action α. When $p = 1$ (i.e. each agent observes a single action), then any mixed action is consistent with the strategy distribution. When $p < 1$ (i.e., each agents observes on average less than one action), then α is the unique consistent outcome.

Structure The next section presents a motivating example. The model is described in Section 3, and Section 1 presents the result. Technical parts of the proof appear in the appendix.

2 Motivating Example

Consider a population in which agents are randomly matched to play the Rock-paper-scissors game, in which each player has three pure actions (*rock*, *paper*, *scissors*), and each action is the unique best reply to the previous action (modulo 3). In each period everyone in the population is matched to play with someone else. Assume that each agent plays the mixed action α at the first round (round 1). At each other round, each agent with a probability of p observes the partner's action in the last round and best replies to it; with the remaining probability of $1 - p$ each agent observes nothing and plays a mixed action β.

What will be the long run behavior of the population? If $p = 1$ it is immediate that the population's behavior will cycle "around" permutations of the initial behavior (as is common in evolutionary models of rock-paper-scissors, see, e.g., the analysis in Cason *et al.*, 2014). Formally, let $n \in \{01, 2, ...\}$:

1. At round $3 \cdot n + 1$ players play *rock* with a probability of $\alpha(rock)$, *paper* with a probability of $\alpha(paper)$, and *scissors* with a probability of $\alpha(scissors)$.

2. At round $3 \cdot n + 2$ players play *rock* with a probability of $\alpha(scissors)$, *paper* with a probability of $\alpha(rock)$, and *scissors* with a probability of $\alpha(paper)$.

3. At round $3 \cdot n + 3$ players play *rock* with a probability of $\alpha(paper)$, *paper* with a probability of $\alpha(scissors)$, and *scissors* with a probability of $\alpha(rock)$.

[2] A key difference between our analysis and the stochastic stability approach to evolutionary analysis (as pioneered by Young, 1993; Kandori *et al.*, 1993) is that in our model a player observes a signal about his partner's past behavior, while in most of the existing literature on stochastic stability the agent observes a signal about the past aggregate behavior in the population.

However, when $p<1$, one can show that the population converges to the following unique behavior (regardless of the initial behavior α):

$$\Pr(rock)=\frac{\beta(rock)+p\cdot\beta(scissors)+p^2\cdot\beta(paper)}{1+p+p^2},$$

$$\Pr(scissors)=\frac{\beta(scissors)+p\cdot\beta(paper)+p^2\cdot\beta(rock)}{1+p+p^2},$$

$$\Pr(paper)=\frac{\beta(paper)+p\cdot\beta(rock)+p^2\cdot\beta(scissors)}{1+p+p^2}.$$

Note that when p is close to one, the unique behavior is close to the uniform mixed profile that assigns a probability of $\frac{1}{3}$ for each action.

Our main result formalizes and extends this example. It shows that when each player observes on average less than one past action of the partner, then the aggregate behavior converges to a unique outcome regardless of the initial conditions, while if players observe on average more than one action, then the long run behavior may depend on the initial condition (and may cycle rather then converge).

3 Model

3.1 Actions, Strategies and Outcomes

We present a reduced form static analysis of a dynamic evolutionarily process of cultural learning (or, alternatively, of a biological evolutionary process) in a large population of agents. The agents in the population are randomly matched into pairs and play a symmetric one-shot game, in which each player has to choose an action $a\in A$.[3] Let $\Delta(A)$ denote the set of mixed actions (distributions over A). We use the letter a (α) to denote a typical pure (mixed) action. With slight abuse of notation let $a\in A$ also denote the element in $\Delta(A)$, which assigns probability 1 to a. We adopt this convention for all probability distributions throughout the paper.

An *observation function* $p\in\Delta(\mathbb{N}\times\mathbb{N})$ is a distribution (with a finite support) over pairs of non-negative integers. Before playing the game, each player privately observes with probability $p(k_1,k_2)$ k_1 independent realized actions of his partner and k_2 realized action profiles played by his partner (the first action in each such profile) and her opponents. Let $C^+(p)=C(p)\setminus\{(0,0)\}$ denote the support of p except $(0,0)$.

Given pair $(k_1,k_2)\neq(0,0)$ let M_{k_1,k_2} denote the set of signals (or messages) that include k_1 observed actions and k_2 observed action profiles: $M_{k_1,k_2}=A^{k_1}\times(A\times A)^{k_2}$. Let $M=M_p$ denote the set of all possible *signals* given observation function p, i.e. $M=\left\{\cup_{(k_1,k_2)\in C(p)}M_{k_1,k_2}\right\}$. Let m denote a typical message (i.e., an element of M). Let $\phi\equiv A^{\emptyset}\times(A\times A)^{\emptyset}$ denote the non-informative signal in which no actions are observed (which occurs with probability $p(0,0)$).

A *strategy* is a mapping $s:M\to\triangle(A)$ that assigns a mixed action to each possible message. Let $s_m\in\triangle(A)$ denote the mixed action played by strategy s after observing signal m. I.e., for each action $a\in A$, $s_m(a)=s(m)(a)$ is the probability that a player who follows strategy s plays action a after observing message m.

Let S denote the set of all strategies, and let $\Sigma\equiv\triangle(S)$ denote the set of finite support distributions over the set of strategies. An element $\sigma\in\Sigma$ is called a *strategy distribution*. Let $\sigma(s)$ denote the probability that

[3] The assumption that the underlying game G is symmetric is essentially without loss of generality (if the game is played within a single population). Asymmetric games can be symmetrized by considering an extended game in which agents are randomly assigned to the different player positions with equal probability, and strategies condition on the assigned role (see, e.g., Selten, 1980).

strategy distribution σ assigns to strategy s. Given a strategy distribution $\sigma \in \Sigma$, let $C(\sigma)$ denote its support (i.e., the set of strategies such that $\sigma(s) > 0$). We interpret $\sigma \in \Sigma$ as representing a population in which $|C(\sigma)|$ strategies coexist, each agent is endowed with one of these strategies according to the distribution of σ. When $|C(\sigma)| = 1$, we identify the strategy distribution with the unique strategy in its support (i.e., $\sigma \equiv s$), in line with the convention adopted above.

Given a finite set of strategies $\tilde{S} \subset S$, an *outcome* $\eta : \tilde{S} \times \tilde{S} \to \triangle(A)$ is a mapping that assigns to each pair of strategies $s, s' \in \tilde{S}$ a mixed action $\eta_s(s')$, which is interpreted as the mixed action played by an agent with strategy s conditional on being matched with a partner with strategy s'. Let $O_{\tilde{S}} \equiv (\triangle(A))^{(\tilde{S} \times \tilde{S})}$ denote the set of all outcomes defined over the set of strategies $\tilde{S}$. The strategy distribution and the outcome together determine the payoffs earned by each agent in the population.

We now present a few definitions for a given strategy distribution $\sigma \in \Sigma$, an outcome $\eta \in O_{C(\sigma)}$, and a strategy $s \in C(\sigma)$. Let $\eta_{s,\sigma} \in \triangle(A)$ be the mixed action played by an agent with strategy s when being matched with a random partner sampled from σ. Formally, for each action $a \in A$:

$$\eta_{s,\sigma}(a) = \sum_{s' \in C(\sigma)} \sigma(s') \cdot \eta_s(s')(a).$$

Let $\psi_{s,\sigma,\eta} \in \triangle(A \times A)$ be the (possibly correlated) mixed action profile that is played when an agent with strategy s is matched with a random opponent sampled from σ. Formally, for each action profile $(a, a') \in A \times A$, where a is interpreted as the action of the agent with strategy s, and a' is interpreted as the action of his partner:

$$\psi_{s,\sigma,\eta}(a, a') = \sum_{s' \in C(\sigma)} \sigma(s') \cdot \eta_s(s')(a) \cdot \eta_{s'}(s)(a').$$

Given a profile of k_1 actions and k_2 action-profiles, $m_{k_1,k_2} = \left((a_i)_{1 \le i \le k_1}, (a_i, a'_i)_{k_1 < i \le k_2}\right) \in M_{k_1,k_2}$, let $\nu_{s,\sigma,\eta}(m_{k_1,k_2})$ denote the probability that a profile of independent observations that include $k_1$realized actions of strategy s and k_2 realized action profiles of strategy s and a random opponent is equal to m_{k_1,k_2}:

$$\nu_{s,\sigma,\eta}(m_{k_1,k_2}) = \prod_{1 \le i \le k_1} \eta_{s',\sigma}(a_i) \cdot \prod_{k_1 < i \le k_2} \psi_{s',\sigma,\eta}(a_i, a'_i),$$

and $\forall (k_1, k_2) \in C^+(p)$, let $\nu_{s,\sigma,\eta}(k_1, k_2) \in \Delta(M_{k_1,k_2})$ be the distribution over signals in M_{k_1,k_2}.

3.2 Consistent Outcomes

Fix an observation function p. When individuals are drawn to play the game their actions are determined by their strategy and the signals they observe. Suppose that the observed signals are sampled from the outcome η and the players play according to the strategy distribution σ, and this induces a new outcome. We require outcomes to be consistent with the strategy distribution in the sense that they generate observations that induces the current outcome to persist. Formally, given a strategy distribution $\sigma \in \Sigma$, let $f_\sigma : O_{C(\sigma)} \to O_{C(\sigma)}$ be the mapping between outcomes that is induced by σ.

$$(f_\sigma(\eta))_s(s')(a) = p(0,0) \cdot s(\phi)(a) +$$
$$\textstyle\sum_{(k_1,k_2) \in C^+(p)} p(k_1, k_2) \cdot \sum_{m_{k_1,k_2} \in M_{k_1,k_2}} \nu_{s,\sigma,\eta}(m_{k_1,k_2}) \cdot s(m_{k_1,k_2})(a)$$

An outcome $\eta \in O_{C(\sigma)}$ is consistent with strategy distribution σ if it is a fixed point of this mapping: $f_\sigma(\eta) \equiv \eta$. The following standard lemma shows that each strategy distribution admits consistent outcomes.

Lemma 1. *For each strategy distribution $\sigma \in \Sigma$ there exists a consistent outcome η.*

Proof. Observe that the space $O_{C(\sigma)}$ is a convex and compact subset of a Euclidean space, and that the mapping $f_\sigma : O_{C(\sigma)} \to O_{C(\sigma)}$ is continuous. Brouwer's fixed-point theorem implies that the mapping σ has a fixed point $\eta*$, which is a consistent outcome by definition. □

The following example shows there are strategy distributions that admit multiple consistent outcomes (case 1) and strategy distributions that admit a unique consistent outcome (case 2).

Example 1. Assume that each player observes a single action with probability p_1 and no actions with probability $1-p_1$ (i.e., $p(1,0)=p_1$, $p(0,0)=1-p_1$). Let $\tilde{s}$ be the following "tit-for-tat" strategy: $\tilde{s}(a)=a$ for each $a \in A$ (i.e., an individual that follows this strategy plays the observed past action of his opponent), and $s(\phi)=\alpha$ for some arbitrary mixed action α. Note that: (1) if $p_1=1$, then any outcome $\eta \in O_{\tilde{s}}$ is consistent with the strategy distribution that assigns mass 1 to strategy $\tilde{s}$; and (2) if $p_1<1$, then one can show that the unique consistent outcome is α.

4 Result

Theorem 1 characterizes which observation structures induce unique consistent outcomes; that is, structures in which every strategy distribution admits a unique consistent outcome. It turns out that an observation structure induces unique consistent outcomes if and only if the expected number of actions that each agent observes about his partner in each round is at most one.

Definition 1. Given function p, let $\mathbf{E}(p)$ denote the expected number of actions observed by each agent before playing the game:

$$\mathbf{E}(p) = \sum_{(k_1,k_2)\in C(p)} p(k_1,k_2)\cdot(k_1+2\cdot k_2).$$

Theorem 1. *The following conditions are equivalent: (1) Every strategy distribution in environment $E=(G,p)$ admits a unique consistent outcome; and (2) $\mathbf{E}(p)\leq 1$ and $p(1,0)<1$.*

Proof. We begin by proving that "¬2" implies "¬1". Assume that $\mathbf{E}(p)>1$.[4] Let a and a' be different actions ($a \neq a'$). Let s^* be the following strategy: play a if the observed actions include a, and play a' otherwise. Consider the strategy distribution in which all agents follow strategy s^*. Consider the outcome η_x that assigns probability $0 \leq x \leq 1$ to action a and the remaining probability $(1-x)$ to action a'. Note that outcome η_x is consistent with s^* if and only if

$$x = Pr\,(\text{observing } a) = \sum_{(k_1,k_2)\in C^+(p)} p(k_1,k_2)\cdot \sum_{m_{k_1,k_2}\in M_{k_1,k_2}} \left(1-(1-x)^{(k_1+2\cdot k_2)}\right).$$

[4]Note that not-2 implies $\mathbf{E}(p)>1$ or $p(1,0)=1$, and $p(1,0)=1$ implies $\mathbf{E}(p)=1$. Recall that the case of $p(1,0)=\mathbf{E}(p)=1$ was dealt with in Example 1 with multiple consistent outcomes.

It is immediate that $x = 0$ always solves this equation, and thus η_0 is a consistent outcome. Next, note that when $x > 0$ is close to 0 the RHS can be (Taylor-)approximated by:

$$\left(\sum_{(k_1,k_2) \in C^+(p)} p\left(k_1, k_2\right) \cdot \sum_{m_{k_1,k_2} \in M_{k_1,k_2}} \left(\left(k_1 + 2 \cdot k_2\right)\right) \cdot x \right) = \mathbf{E}\left(p\right) \cdot x > x.$$

For $x = 1$ the RHS is $\sum_{(k_1,k_2) \in C^+(p)} p\left(k_1, k_2\right) \leq 1$, so if $\sum_{(k_1,k_2) \in C^+(p)} p\left(k_1, k_2\right) = 1$ then $x = 1$ is also a solution and if $\sum_{(k_1,k_2) \in C^+(p)} p\left(k_1, k_2\right) \leq 1 < 1$ then by continuity of the RHS, there is some $x \in (0, 1)$ that solves the equation. Thus there is $\eta_x \neq \eta_0$ that is also a consistent outcome of s^*.

In what follows, we sketch the proof of the opposite direction: "2" implies "1", while leaving some formal details about the rigorous definition of the various norms to Appendix A. Let σ be an arbitrary strategy distribution, and let η and η' be two outcomes. Under an appropriate choice of norm, the distance between the outcomes $f_\sigma\left(\eta\right)$ and $f_\sigma\left(\eta'\right)$ is bounded by the expected distance between the distributions of signals:

$$\left\| f_\sigma\left(\eta\right) - f_\sigma\left(\eta'\right) \right\| \leq \sum_{k_1,k_2 \in C^+(p)} p\left(k_1, k_2\right) \cdot \left\| \nu_{s,\sigma,\eta}\left(k_1, k_2\right) - \nu_{s,\sigma,\eta'}\left(k_1, k_2\right) \right\|.$$

This is because the mapping f_σ can induce different outcomes only to the extent that the observed signals were different. Next it can be shown that the distance between the signal distributions is bounded by the length of the signal times the distance between the outcomes:

$$\left\| \nu_{s,\sigma,\eta}\left(k_1, k_2\right) - \nu_{s,\sigma,\eta'}\left(k_1, k_2\right) \right\| \leq \left(k_1 + 2 \cdot k_2\right) \cdot \left\| \eta - \eta' \right\|.$$

This is because two observed actions differ with a probability of at most $\|\eta - \eta'\|$, and the probability of $k_1 + 2 \cdot k_2$ observed actions to differ in at least one action is at most $k_1 + 2 \cdot k_2$ times $\|\eta - \eta'\|$ (with strict inequality if $(k_1, k_2) \neq (1, 0)$). Substituting the second inequality in the first one yields:

$$\left\| f_\sigma\left(\eta\right) - f_\sigma\left(\eta'\right) \right\| \leq \sum_{k_1,k_2 \in C^+(p)} p\left(k_1, k_2\right) \cdot \left\| \eta - \eta' \right\| = \mathbf{E}\left(p\right) \cdot \left\| \eta - \eta' \right\|. \tag{1}$$

This implies that if $\mathbf{E}\left(p\right) < 1$ (or if $\mathbf{E}\left(p\right) = 1 < p\left(1, 0\right)$) then f_σ is a contraction mapping: $\left\| f_\sigma\left(\eta\right) - f_\sigma\left(\eta'\right) \right\| < \left\| \eta - \eta' \right\|$, which implies uniqueness. □

A Technical Aspects of the Proof ("2" Implies "1")

Assume that $\mathbf{E}\left(p\right) \leq 1$ and $p\left(1, 0\right) < 1$. We show that every strategy distribution in E admits a unique consistent outcome. Let σ be a strategy distribution, and let $\eta \neq \eta' \in O_{c(\sigma)}$. In order to shorten the notations, we omit the subscript σ in the remainder of the proof. For example we write $\eta_s\left(a\right)$ instead of $\eta_{s,\sigma}\left(a\right)$. In what follows we show that f (i.e. f_σ) is a contraction mapping (which implies admitting a unique consistent outcome).

A.1 Definitions of Norms

We measure distance between (finite support) probability distributions with the L_1-norm as follows: let X be a finite set and $\Delta\left(X\right)$ the set of probability distributions on X. Given distributions $\xi, \xi' \in \Delta\left(X\right)$, their distance is defined as the sum of the absolute differences in the weights they assign to the different elements of X:

$$\left\|\xi-\xi'\right\|_1=\sum_{x\in X}\left|\xi\left(x\right)-\xi'\left(x\right)\right|.$$

We measure distance between profiles of distributions with the help of the L_∞-norm. For any two profiles of distributions $\gamma=\left(\xi_i\right)_{i\in I},\gamma'=\left(\xi'_i\right)_{i\in I}$, we define

$$\left\|\gamma-\gamma'\right\|_\infty \quad=\quad \max_{i\in I}\left\|\xi_i-\xi'_i\right\|_1.$$

Let $\nu_\eta\left(k_1,k_2\right)=\left(\nu_{s,\eta}\left(k_1,k_2\right)\right)_{s\in C(\sigma)}$ denote the profile of distributions over signals in M_{k_1,k_2} for the various strategies in the support of σ. Then, in particular:

$$\left\|\nu_\eta\left(k_1,k_2\right)-\nu_{\eta'}\left(k_1,k_2\right)\right\|_\infty=\max_{s\in C(\sigma)}\left\|\nu_{s,\eta}\left(k_1,k_2\right)-\nu_{s,\eta'}\left(k_1,k_2\right)\right\|_1.$$

Similarly, since η_s can be interpreted as representing the profile set $\left\{\eta_s\left(s'\right)\right\}_{s'\in c(\sigma)}$:

$$\left\|\eta_s-\eta'_s\right\|_\infty=\max_{s'\in C(\sigma)}\left\|\eta_s\left(s'\right)-\eta'_s\left(s'\right)\right\|_1.$$

Finally, we use two norms $\|\|_{\infty,\infty}$ and $\|\|_{\infty,1}$ to measure distances between outcomes η and η':

$$\left\|\eta-\eta'\right\|_{\infty,\infty}=\max_{s\in C(\sigma)}\left\|\eta_s-\eta'_s\right\|_\infty,\quad \left\|\eta-\eta'\right\|_{\infty,1}=\max_{s\in C(\sigma)}\left\|\eta_s-\eta'_s\right\|_1.$$

Note that $\left\|\eta_s-\eta'_s\right\|_1\leq\left\|\eta_s-\eta'_s\right\|_\infty$ and $\left\|\eta-\eta'\right\|_{\infty,1}\leq\left\|\eta-\eta'\right\|_{\infty,\infty}$.

A.2 Bounding the Distance Between Combinations of Actions and Action Profiles

Let $k_1,k_2\geq 1$. We begin by showing that the distance between the distribution of messages in M_{k_1,k_2} is at most the sum of the distances between the distributions in $M_{k_1}\equiv M_{k_1,0}$ and the distributions in $M_{k_2}\equiv M_{0,k_2}$. For $m_{k_i}\in M_{k_i}$, $i=1,2$, we write $\nu_{s,\eta}\left(m_{k_i}\right)$ instead of $\nu_{s,\eta}\left(m_{k_1,k_2}\right)$.

$$\begin{aligned}
&\left\|\nu_{s,\eta}\left(k_1,k_2\right)-\nu_{s,\eta'}\left(k_1,k_2\right)\right\|_1=\sum_{\left(m_{k_1,k_2}\right)\in M_{k_1,k_2}}\left|\nu_{s,\eta}\left(m_{k_1,k_2}\right)-\nu_{s,\eta'}\left(m_{k_1,k_2}\right)\right|\\
&=\sum_{m_{k_1}\in M_{k_1}}\sum_{m_{k_2}\in M_{k_2}}\left|\nu_{s,\eta}\left(m_{k_1}\right)\cdot\nu_{s,\eta}\left(m_{k_2}\right)-\nu_{s,\eta'}\left(m_{k_1}\right)\cdot\nu_{s,\eta'}\left(m_{k_2}\right)\right| \qquad (2)\\
&=\sum_{m_{k_1}\in M_{k_1}}\sum_{m_{k_2}\in M_{k_2}}\left|\nu_{s,\eta}\left(m_{k_1}\right)\cdot\left(\nu_{s,\eta}\left(m_{k_2}\right)-\nu_{s,\eta'}\left(m_{k_2}\right)\right)+\nu_{s,\eta'}\left(m_{k_2}\right)\cdot\left(\nu_{s,\eta}\left(m_{k_1}\right)-\nu_{s,\eta'}\left(m_{k_1}\right)\right)\right|\\
&<\sum_{m_{k_1}\in M_{k_1}}\sum_{m_{k_2}\in M_{k_2}}\left|\nu_{s,\eta}\left(m_{k_1}\right)\cdot\left(\nu_{s,\eta}\left(m_{k_2}\right)-\nu_{s,\eta'}\left(m_{k_2}\right)\right)\right|+\left|\nu_{s,\eta'}\left(m_{k_2}\right)\cdot\left(\nu_{s,\eta}\left(m_{k_1}\right)-\nu_{s,\eta'}\left(m_{k_1}\right)\right)\right|\\
&=\sum_{m_{k_1}\in M_{k_1}}\sum_{m_{k_2}\in M_{k_2}}\nu_{s,\eta}\left(m_{k_1}\right)\cdot\left|\left(\nu_{s,\eta}\left(m_{k_2}\right)-\nu_{s,\eta'}\left(m_{k_2}\right)\right)\right|+\nu_{s,\eta'}\left(m_{k_2}\right)\cdot\left|\left(\nu_{s,\eta}\left(m_{k_1}\right)-\nu_{s,\eta'}\left(m_{k_1}\right)\right)\right|\\
&=\sum_{m_{k_2}\in M_{k_2}}\left|\left(\nu_{s,\eta}\left(m_{k_2}\right)-\nu_{s,\eta'}\left(m_{k_2}\right)\right)\right|+\sum_{m_{k_1}\in M_{k_1}}\left|\left(\nu_{s,\eta}\left(m_{k_1}\right)-\nu_{s,\eta'}\left(m_{k_1}\right)\right)\right| \qquad (3)\\
&=\left\|\nu_{s,\eta}\left(k_1,0\right)-\nu_{s,\eta'}\left(k_1,0\right)\right\|_1+\left\|\nu_{s,\eta}\left(0,k_2\right)-\nu_{s,\eta'}\left(0,k_2\right)\right\|_1.
\end{aligned}$$

Equality (2) is due to the independence of different observations. The next equality is derived by adding and subtracting $\nu_{s,\eta}(m_{k_1}) \cdot \nu_{s,\eta'}(m_{k_2})$. The inequality is strict because the elements inside the "$||$" in the l.h.s. of the inequality include both positive and negative numbers. Eq. (3) holds because each sum adds the probabilities of disjoint and exhausting events.

A.3 Bounding the Distance Between Action Profiles

We show that the distance ($\|\|_1$) between two distributions over observed action profiles is at most twice the distance ($\|\|_{\infty,\infty}$) between the corresponding outcomes.

$$\begin{aligned}
&\|\psi_{s,\eta} - \psi_{s,\eta'}\|_1 = \sum_{(a,a')\in A^2} |\psi_{s,\eta}(a,a') - \psi_{s,\eta'}(a,a')| \\
&= \sum_{(a,a')\in A^2} \left| \sum_{s'\in C(\sigma)} \sigma(s') \cdot (\eta_s(s')(a) \cdot \eta_{s'}(s)(a') - \eta'_s(s')(a) \cdot \eta'_{s'}(s)(a')) \right| \\
&\leq \sum_{(a,a')\in A^2} \sum_{s'\in C(\sigma)} \sigma(s') \left|(\eta_s(s')(a) \cdot \eta_{s'}(s)(a') - \eta'_s(s')(a) \cdot \eta'_{s'}(s)(a'))\right| \\
&= \sum_{s'\in C(\sigma)} \sigma(s') \cdot \sum_{(a,a')\in A^2} \left|(\eta_s(s')(a) \cdot \eta_{s'}(s)(a') - \eta'_s(s')(a) \cdot \eta'_{s'}(s)(a'))\right| \\
&= \sum_{s'\in C(\sigma)} \sigma(s') \cdot \sum_{(a,a')\in A^2} \left|(\eta_s(s')(a) \cdot (\eta_{s'}(s)(a') - \eta'_{s'}(s)(a')) + \eta'_{s'}(s)(a')(\eta_s(s')(a) - \eta'_{s'}(s')(a)))\right| \\
&< \sum_{s'\in C(\sigma)} \sigma(s') \cdot \sum_{(a,a')\in A^2} (\eta_s(s')(a) \cdot |\eta_{s'}(s)(a') - \eta'_{s'}(s)(a')| + \eta'_{s'}(s)(a') \cdot |\eta_s(s')(a) - \eta'_{s'}(s')(a)|) \\
&= \sum_{s'\in C(\sigma)} \sigma(s') \cdot \sum_{a'\in A} |\eta_{s'}(s)(a') - \eta'_{s'}(s)(a')| + \sum_{a\in A} |\eta_s(s')(a) - \eta'_{s'}(s')(a)| \\
&= \sum_{s'\in C(\sigma)} \sigma(s') \cdot (\|\eta_{s'}(s) - \eta'_{s'}(s)\|_1 + \|\eta_s(s') - \eta'_{s'}(s')\|_1) \\
&\leq \sum_{s'\in C(\sigma)} \sigma(s') \cdot (\|\eta_{s'} - \eta'_{s'}\|_\infty + \|\eta_s - \eta'_s\|_\infty) \\
&\leq \sum_{s'\in C(\sigma)} \sigma(s') \cdot \left(2 \cdot \|\eta - \eta'\|_{\infty,\infty}\right) = 2 \cdot \|\eta - \eta'\|_{\infty,\infty}.
\end{aligned}$$

The second inequality is strict because the elements inside the "$||$" in the l.h.s. of the strict inequality include both positive and negative elements.

A.4 Bounding the Distance Between Sequences of Actions

Let $k \geq 2$. Next we show that the distance ($\|\|_1$) between two sequences of k observed actions is at most k times the distance ($\|\|_{\infty,\infty}$) between the corresponding outcomes:

$$\|\nu_{s,\eta}(k,0) - \nu_{s,\eta'}(k,0)\|_1 = \sum_{(a_i)_{i\leq k_1}\in A^k} \left|\nu_{s,\eta}\left((a_i)_{i\leq k}\right) - \nu_{s,\eta'}\left((a_i)_{i\leq k}\right)\right| =$$

$$= \sum_{(a_i)_{i\le k_1}\in A^k} \left| \prod_{i\le k} \eta_s(a_i) - \prod_{i\le k} \eta'_s(a_i) \right| = \sum_{(a_i)_{i\le k}\in A^k} \left| \sum_{i\le k} (\eta_s(a_i) - \eta'_s(a_i)) \cdot \prod_{j>i} \eta_s(a_j) \cdot \prod_{j<i} \eta'_s(a_j) \right| \quad (4)$$

$$< \sum_{(a_i)_{i\le k}\in A^k} \sum_{i\le k} \left(|\eta_s(a_i) - \eta'_s(a_i)| \cdot \prod_{j>i} \eta_s(a_j) \cdot \prod_{j<i} \eta'_s(a_j) \right)$$

$$= \sum_{i\le k} \left(\sum_{a\in A} |\eta_s(a) - \eta'_s(a)| \cdot \left(\sum_{(a_j)_{j>i}\in A^{n-i}} \prod_{j>i} \eta_s(a_j) \right) \cdot \left(\sum_{(a_j)_{j<i}\in A^{i-1}} \prod_{j<i} \eta'_s(a_j) \right) \right)$$

$$= \sum_{i\le k} \left(\sum_{a\in A} |\eta_s(a) - \eta'_s(a)| \cdot 1 \cdot 1 \right) \quad (5)$$

$$= k \cdot \sum_{a\in A} |\eta_s(a) - \eta'_s(a)| = k \cdot \|\eta_s - \eta'{}_s\|_1 \le k \cdot \|\eta - \eta'\|_{\infty,1} \le k \cdot \|\eta - \eta'\|_{\infty,\infty}.$$

The first equality in Eq. (4) is due to the independence of different observations, and the second equality is implied by adding to the sum elements that cancel out (appearing once with a positive sign and once with a negative sign). The inequality is strict because the set of numbers inside the "||" in the l.h.s. of the inequality include both positive and negative elements. Equality (5) holds because each sum adds the probabilities of disjoint and exhausting events.

An analogous argument yields the same result for observed tuples of action profiles (where the strict inequality is implied by Section A.3): $\|\nu_{s,\eta}(0,k) - \nu_{s,\eta'}(0,k)\|_1 \le k \cdot \|\psi_{s,\eta} - \psi_{s,\eta'}\|_1 < k \cdot 2 \cdot \|\eta - \eta'\|_{\infty,\infty}$.

A.5 Showing that $f(\eta)$ is a Contraction Mapping

We bound the distance between $(f(\eta))_s(s')$ and $(f(\eta'))_s(s')$ as follows.

$$\|(f(\eta))_s(s') - (f(\eta'))_s(s')\|_1 = \sum_{a\in A} |(f(\eta))_s(s')(a) - (f(\eta'))_s(s')(a)|$$

$$= \sum_{a\in A} \sum_{(k_1,k_2)\in C^+(f)} p(k_1,k_2) \cdot \left| \sum_{m_{k_1,k_2}\in M_{k_1,k_2}} (\nu_{s,\eta}(m_{k_1,k_2}) - \nu_{s,\eta'}(m_{k_1,k_2})) \cdot s(m_{k_1,k_2})(a) \right|$$

$$\le \sum_{a\in A} \sum_{(k_1,k_2)\in C^+(f)} p(k_1,k_2) \cdot \left| \sum_{m_{k_1,k_2}\in M_{k_1,k_2}} (\nu_{s,\eta}(m_{k_1,k_2}) - \nu_{s,\eta'}(m_{k_1,k_2})) \right| \quad (6)$$

$$= \sum_{a\in A} \sum_{(k_1,k_2)\in C^+(f)} p(k_1,k_2) \cdot \|\nu_{s,\eta}(k_1,k_2) - \nu_{s,\eta'}(k_1,k_2)\|_1$$

$$\le \sum_{a\in A} \sum_{(k_1,k_2)\in C^+(f)} p(k_1,k_2) \cdot \left(\|\nu_{s,\eta}(k_1,0) - \nu_{s,\eta'}(k_1,0)\|_1 + \|\nu_{s,\eta}(0,k_2) - \nu_{s,\eta'}(0,k_2)\|_1 \right) \quad (7)$$

$$\le \sum_{a\in A} \sum_{(k_1,k_2)\in C^+(f)} p(k_1,k_2) \cdot \left(k_1 \cdot \|\eta - \eta'\|_{\infty,\infty} + k_2 \cdot 2 \cdot \|\eta - \eta'\|_{\infty,\infty} \right) \quad (8)$$

$$= \sum_{a\in A} \|\eta - \eta'\|_{\infty,\infty} \cdot \sum_{(k_1,k_2)\in C^+(f)} p(k_1,k_2) \cdot (k_1 + k_2 \cdot 2) = \sum_{a\in A} \|\eta - \eta'\|_{\infty,\infty} \cdot \mathbf{E}(p) \le \sum_{a\in A} \|\eta - \eta'\|_{\infty,\infty}.$$

Inequality (6) is implied by omitting the term $0 \leq s\left(m_{k_1,k_2}\right)(a) \leq 1$. Inequality (7) is derived by result of Section A.2. Inequality (8) is implied by Section A.4 (with a strict inequality if $p\left(k_1,k_2\right) > 0$ for any $k_1 \geq 2$ or $k_2 \geq 1$). The last inequality is strict if $\mathbf{E}(p) < 1$. Thus, at least one of these inequalities is strict if $\mathbf{E}(p) \leq 1$ and $p(1,0) < 1$. Therefore, we obtain the following strict inequality (which implies that f is a contraction mapping):

$$\left\| f(\eta) - f(\eta') \right\|_{\infty,\infty} \leq \max_{s,s' \in C(\sigma)} \left\| \left(f_\sigma(\eta)\right)_s(s') - \left(f_\sigma(\eta')\right)_s(s') \right\|_1 \quad < \quad \left\| \eta - \eta' \right\|_{\infty,\infty}.$$

References

Bhaskar, V, Mailath, George J, & Morris, Stephen. 2013. A foundation for Markov equilibria in sequential games with finite social memory. *The Review of Economic Studies*, **80**(3), 925–948.

Cason, Timothy N, Friedman, Daniel, & Hopkins, ED. 2014. Cycles and Instability in a Rock–Paper–Scissors Population Game: A Continuous Time Experiment. *The Review of economic studies*, **81**(1), 112–136.

Heller, Yuval, & Mohlin, Erik. 2015. *Observations on cooperation.*

Kandori, Michihiro, Mailath, George J, & Rob, Rafael. 1993. Learning, mutation, and long run equilibria in games. *Econometrica*, 29–56.

Nowak, Martin A, & Sigmund, Karl. 1998. Evolution of indirect reciprocity by image scoring. *Nature*, **393**(6685), 573–577.

Rosenthal, Robert W. 1979. Sequences of games with varying opponents. *Econometrica*, 1353–1366.

Selten, Reinhard. 1980. A note on evolutionarily stable strategies in asymmetric animal conflicts. *Journal of Theoretical Biology*, **84**(1), 93–101.

Young, H Peyton. 1993. The evolution of conventions. *Econometrica*, 57–84.