Instability of Defection in the Prisoner's Dilemma: Best Experienced Payoff Dynamics Analysis

Arigapudi, Srinivas and Heller, Yuval and Milchtaich, Igal

Bar-Ilan University, University of Wisconsin, Bar-Ilan University

11 April 2020

Instability of Defection in the Prisoner's Dilemma: Best Experienced Payoff Dynamics Analysis*

Srinivas Arigapudi†, Yuval Heller‡, Igal Milchtaich§

April 12, 2020

Abstract

We study population dynamics under which each revising agent tests each strategy k times, with each trial being against a newly drawn opponent, and chooses the strategy whose mean payoff was highest. When $k = 1$, defection is globally stable in the prisoner's dilemma. By contrast, when $k > 1$ we show that there exists a globally stable state in which agents cooperate with probability between 28% and 50%. Next, we characterize stability of strict equilibria in general games. Our results demonstrate that the empirically-plausible case of $k > 1$ can yield qualitatively different predictions than the case of $k = 1$ that is commonly studied in the literature.

Keywords: learning, cooperation, best experienced payoff dynamics, sampling equilibrium, evolutionary stability. **JEL codes:** C72, C73.

1. Introduction

The standard approach in game theory assumes that players play a Nash equilibrium. However, in some environments where the players have limited information about the strategic situation, Nash equilibrium prediction is hard to justify. Consider the following example from Osborne and Rubinstein (1998). You are new to town and are planning your route to work. How do you decide which road to take? You know that other people

*We wish to thank Luis R. Izquierdo, Segismundo S. Izquierdo, Ron Peretz and William Sandholm for many helpful discussions. SA gratefully acknowledges the financial support from Sandwich research fellowship by the Israeli Council of Higher Education and Bar-Ilan University.

†Department of Economics, University of Wisconsin, 1180 Observatory Drive, Madison, WI 53706. e-mail: arigapudi@ssc.wisc.edu; website: sites.google.com/view/srinivasarigapudi/home.

‡Department of Economics, Bar-Ilan University, Ramat Gan 5290002, Israel. e-mail: yuval.heller@biu.ac.il; website: https://sites.google.com/site/yuval26/.

§Department of Economics, Bar-Ilan University, Ramat Gan 5290002, Israel. e-mail: igal.milchtaich@biu.ac.il; website: https://faculty.biu.ac.il/~milchti/.

use the roads, but have no idea which road is most congested. One plausible procedure is to try each route several times and then permanently adopt the one that was (on average) best. The outcome of this procedure is stochastic: you may sample the route that is in fact the best on a day when a baseball game congests it. Once you select your route, you become part of the environment that determines other drivers' choices.

This procedure is formalized as follows. Consider agents in a large population who are randomly matched to play a symmetric game with a finite set of actions. Agents occasionally revise their action (which can be interpreted as agents occasionally leaving the population, and being replaced by new agents). Each revising agent samples each feasible action k times and chooses the action that yields the highest average payoff (with an arbitrary tie-breaking rule).

This procedure induces a dynamic process according to which the distribution of actions in the population evolves (best experienced payoff dynamics, Sethi, 2000; Sandholm et al., 2019). An $S(k)$ equilibrium is a rest point of the dynamics. An $S(k)$ equilibrium α^* is locally stable if any sufficiently close initial population converges into playing α^*, and it is (almost) globally stable if any interior initial population converges into playing α^*.

The existing literature on payoff sampling equilibria (as surveyed below) has mainly focused on $S(1)$ equilibria, due to their tractability. It seems plausible that real life behavior would rely on sampling each action more than once. A key insight of our analysis is that sampling actions a couple of times might lead to qualitatively different results than sampling each action only once. In particular, in the prisoner's dilemma game, $S(1)$ dynamics yield the Nash equilibrium behavior, while $S(k)$ dynamics (for small k's larger than one) induce substantial cooperation.

Recall that each player in the prisoner's dilemma game has two actions, cooperation c and defection d, and the payoffs are as in Table 1, where $g, l > 0$. Sethi (2000) has shown that defection is the unique globally stable $S(1)$ equilibrium. Our first result shows that for any $k \geq 2$ for which the gains from defection are not too large (specifically, $g, l < \frac{1}{k-1}$), the game admits a globally stable state in which the rate of cooperation is between 28% and 50%.

	c	d
c	1 , 1	-l , 1+g
d	1+g , -l	0 , 0

Table 1: Prisoner's Dilemma Payoff Matrix

Our second result analyzes the local stability of the state in which everyone defects (without restricting the parameters g and l). It shows that defection is locally stable iff $k = 1$

or $l > \frac{1}{k-1}$. Our final result extends the analysis of local stability to any strict equilibrium of any generic symmetric game. It presents a simple necessary and sufficient condition for a strict equilibrium (a^*, a^*) to be $S(k)$ locally stable (tightening the conditions presented in Sethi, 2000; Sandholm et al., 2020). Roughly speaking, the condition is that in any set of actions A' that does not include a^* there is an action that never yields the highest payoff when the corresponding sample includes a single occurrence of an action in A' and all the other sampled actions are a^*.

The predictions of our model match quite well the stylized facts about the behavior of subjects playing one-shot prisoner's dilemma. The meta-study of (Mengel, 2018, Tables A.3, B.5) summarizes 29 sessions of lab experiments of that game in a "stranger" setting from 16 papers (with various values of g and l with median one). The average rate of cooperation in these experiments is 37%, and this rate is decreasing in l but is independent of g.[1]

Our predictions might have even a better fit to experiments in which subjects have only partial information about the payoff matrix (a setting which might be relevant to many real-life interactions), such as a "black box" setting in which players do not know the game's structure, and they only observe their realized payoffs (see, e.g., Nax and Perc, 2015; Nax et al., 2016; Burton-Chellew et al., 2017).

Outline: Section 1.1 reviews the related literature. In Section 2, we introduce our model and solution concept. We analyze the prisoner's dilemma in Section 3. We characterize the stability of strict equilibria in generic games in Section 4.

1.1 Related Literature

The payoff sampling dynamics approach employed in this paper was pioneered by Osborne and Rubinstein (1998) and Sethi (2000). The approach has been used in a variety of applications, including bounded-rationality models in industrial organization (Spiegler, 2006a,b), coordination games (Ramsza, 2005), trust and delegation of control (Rowthorn and Sethi, 2008), market entry (Chmura and Güth, 2011), ultimatum games (Mikekisz and Ramsza, 2013), common-pool resources (Cárdenas et al., 2015), contributions to public goods (Mantilla et al., 2018), and finitely-repeated games (Sethi, 2019).

Most of these papers mainly focus on $S(1)$ dynamics in which each action is sampled once. Two exceptions are Sandholm et al. (2019), which analyzes the stable $S(k)$ equilibrium in a centipede game and shows that it leads to cooperative behavior even when the number

[1] Analysis of the $S(2)$ dynamics shows a small influence of g. The rate of cooperation in the $S(2)$ globally stable state is 28% when $g, l < 1$ (Theorem 1), 24% when $l < 1 < g$ (result omitted for brevity), and 0% when $l > 1$ (Proposition 2).

of trials k of each strategy is large, and Sandholm et al. (2020), which develops general stability and instability criteria of $S(k)$ equilibria in general classes of games.

A related alternative approach is *action sampling dynamics* (AKA, sample best response dynamics), according to which each revising agent obtains a small random sample of other players' actions, and chooses the action that is a best reply to that sample (see, for example, Sandholm, 2001; Kosfeld et al., 2002; Kreindler and Young, 2013; Oyama et al., 2015; Heller and Mohlin, 2018; Salant and Cherry, 2019). The action sampling approach is a plausible heuristic when the players know the payoff matrix and are capable of strategic thinking but do not know the exact distribution of actions in the population.

2. Model

We consider a unit-mass continuum of agents who are randomly matched to play a symmetric two-player normal form game $G = \{A, u\}$, where $A = \{a_1, a_2, \ldots, a_m\}$ is the (finite) set of actions and $u : A \times A \to \mathbb{R}$ is the payoff function. An agent taking action a_i against an opponent playing a_j receives payoff $u(a_i, a_j)$.

Aggregate behavior in the population is described by a *population state* α lying in the simplex $\Delta \equiv \{\alpha = (\alpha_{a_i})_{i=1}^{m} \in \mathbb{R}_{+}^{m} \mid \sum_{i=1}^{m} \alpha_{a_i} = 1\}$, with α_{a_i} representing the fraction of agents in the population using action a_i. The standard basis vector $e_a \in \Delta$ represents the pure, or monomorphic, state at which all agents play action a. The set of *interior population states*, in which all actions are used by a positive mass of agents, is $Int(\Delta) \equiv \Delta \cap \mathbb{R}_{++}^{m}$.

A sampling procedure involves the testing of the different actions against randomly-drawn opponents. Agents occasionally receive opportunities to switch actions. These opportunities do not depend on the currently used actions. That is, when the population state is $\alpha(t)$, the proportion of agents using an action a among the agents who revise between time t and $t + dt$ is equal to their proportion in the population $\alpha_a(t)$.

When an agent receives a revision opportunity, he tries each of the feasible actions k times, using it each time against a newly drawn opponent from the population. Thus, the probability that the opponent's action is any $a \in A$ is $\alpha_a(t)$. The agent then chooses the action that yielded the highest mean payoff in these trials, employing some tie-breaking rule if more than one action yields the highest mean payoff. All of our results hold for any tie-breaking rule. Denote the probability that the chosen action is a by $w_{a,k}(\alpha(t))$.

As a result of the revision procedure described above, the expected change in the number of agents using an action a during an infinitesimal time interval of duration dt is

$$w_{a,k}(\alpha(t))dt - \alpha_a(t)dt. \tag{2.1}$$

The first term in Eq. (2.1) is an inflow term, representing the expected number of revising agents who switch to action a, while the second term is an outflow term, representing the expected number of revising agents who currently play that action. In the limit $dt \to 0$, we obtain the rate of change of the fraction of agents using a:

$$\dot{\alpha}_a(t) = w_{a,k}(\alpha(t)) - \alpha_a(t). \tag{2.2}$$

This system of differential equation is called the *k-payoff sampling dynamic*. Its rest points are called $S(k)$ equilibria.

Definition 1 (Osborne and Rubinstein, 1998)**.** A population state $\alpha^* \in \Delta$ is an $S(k)$ *equilibrium* if $w_{a,k}(\alpha^*) = \alpha^*_a$ for all $a \in A$.

An equilibrium is (locally) asymptotically stable if a population beginning near it remains close, and eventually converges to the equilibrium.

Definition 2. An $S(k)$ equilibrium α^* is *asymptotically stable* if:

1. (*Lyapunov stability*) for every neighborhood U of α^* in Δ there is a neighborhood $V \subset U$ of α^* such that if $\alpha(0) \in V$, then $\alpha(t) \in U$ for all $t > 0$; and

2. there is some neighborhood U of α^* in Δ such that all trajectories initially in U converge to α^*, that is, $\alpha(0) \in U$ implies $\lim_{t\to\infty} \alpha(t) = \alpha^*$.

An equilibrium is (almost) globally asymptotically stable if the population converges to it from any initial interior state.

Definition 3. An $S(k)$ equilibrium α^* is *globally asymptotically stable* if all interior trajectories converge to α^*, that is, $\alpha(0) \in Int(\Delta)$ implies $\lim_{t\to\infty} \alpha(t) = \alpha^*$.

3. Prisoner's Dilemma

This section focuses on the prisoner's dilemma game. The set of actions is given by $A = \{c, d\}$, where c is interpreted as cooperation and d as defection, and the payoffs are as described in Table 1, with $g, l > 0$.

3.1 $S(1)$ Analysis

Our preliminary result (adapted from Sethi, 2000) shows that defection is globally stable.

Proposition 1. *Everyone defecting is $S(1)$ globally asymptotically stable.*

Proof. When an agent samples the action c (henceforth, the *c-sample*), he obtains payoff 1 with probability α_c and payoff $-l$ with probability α_d. Similarly, when sampling action d (the *d-sample*), the payoff is $1+g$ with probability α_c and 0 with probability α_d. The (mean) payoff in the c-sample is higher than the payoff of the d-sample iff the c-sample yields payoff 1 (which happens with probability α_c) and the d-sample yields payoff 0 (which happens with probability α_d). Thus, the probability of action c being superior is given by $w_{c,1}(\alpha) = \alpha_c\alpha_d$. The 1-payoff sampling dynamic is therefore given by

$$\dot{\alpha}_c = w_{c,1}(\alpha) - \alpha_c = \alpha_c\alpha_d - \alpha_c = \alpha_c(1-\alpha_c) - \alpha_c = -\alpha_c^2 \qquad (< 0 \text{ for } \alpha_c > 0).$$

The unique rest point $\alpha_c^* = 0$ is the unique $S(1)$ equilibrium, and it is easy to see that it is globally asymptotically stable. □

3.2 $S(k)$ Analysis for small g, l

Next we show that for any $k \geq 2$ and sufficiently small g, l (specifically, $g, l < \frac{1}{k-1}$), the prisoner's dilemma game admits a globally asymptotically stable $S(k)$ equilibrium, in which the frequency of cooperation is between 28% and 50% and is increasing in k.

Theorem 1. *Assume that $k \geq 2$ and that $g, l < \frac{1}{k-1}$. Then there exists an $S(k)$ globally asymptotically stable equilibrium α^k that satisfies $0.28 < \alpha_c^k < 0.5$. Moreover, $\alpha_c^{k'} < \alpha_c^k$ for any $1 < k' < k$.*

Proof. The proof is divided into a number of claims. In what follows we state and present a sketch of the proof for each claim. The formal proofs are given in Appendix A.

Notation Let $p \equiv \alpha_c$ denote the proportion of cooperating agents. For $j \leq k$ let $f_{k,p}(j) \equiv \binom{k}{j} p^j (1-p)^{k-j}$ be the probability mass function of a binomial random variable with parameters k and p. Let $Tie(k,p) = \sum_{j=0}^{k} (f_{k,p}(j))^2$ be the probability of having a tie between two independent binomial random variables with parameters k, p, and let $Win(k,p) = 0.5 \cdot (1 - Tie(k,p)$ be the probability of the first random variables having a larger value than the remaining variable.

Claim 1. *Assume that $g, l \in (0, \frac{1}{k-1})$. The k-payoff sampling dynamic is given by*

$$\dot{p} = Win(k,p) - p. \tag{3.1}$$

Sketch of Proof. The condition $g, l < \frac{1}{k-1}$ implies that action c has a higher mean payoff iff the c-sample includes more cooperating opponents than the d-sample does. The number of cooperators in each sample has a binomial distribution with parameters k and p, so the probability of c having a higher mean payoff is $Win(k,p)$ (which we substitute in (2.2)). $\square$

For $k \geq 2$ and $0 \leq p \leq 1$, denote the expression on the right-hand side of Eq. (3.1) by:

$$h_k(p) = Win(k,p) - p. \tag{3.2}$$

Claim 2. *For all $k \geq 2$, the following conditions hold: $h_k(0) \equiv 0, h_k(1) = -1$ and $h'_k(0) > 0$.*

Sketch of Proof. When $p = 0$ (resp., $= 1$), in both samples all the opponents are defectors (resp., cooperators), which implies that $Win(k,p) = 0$ for $p \in \{0,1\}$, which, in turn, implies that $h_k(0) = 0$ and $h_k(1) = -1$. Next, observe that for $p = \epsilon << 1$, $Win(k,p) \approx k\epsilon$, which is approximately the probability of having at least one cooperator in the c-sample. Thus, $h_k(\epsilon) \approx k\epsilon - \epsilon$, which implies that $h'_k(0) = k - 1 > 0$.

$\square$

Claim 3. *For $k \geq 2$, the function $h_k(p)$ is concave in p and for $p \in (0,1)$ we have $h_k(p) < h_{k+1}(p)$. In addition, $h_k\left(\frac{1}{2}\right) < 0$ and $\lim_{k\to\infty} h_k\left(\frac{1}{2}\right) = 0$.*

Sketch of Proof. Observe that $Tie(k,p)$ is close to 1 when p is close either to zero or to one, and that $Tie(k,p)$ becomes smaller in the intermediate range of p-s. The formal proof shows (by analyzing the characteristic function) that $Tie(k,p)$ is (1) convex in p, (2) decreasing in k (i.e., the larger the number of actions in each sample, the smaller the probability of having exactly the same number of cooperations in both samples), and (3) converges to zero as k converges to ∞. This implies that $h_k(p) \equiv (0.5 \cdot (1 - Tie(k,p)) - p$ is concave in p and increasing in k, and that

$$h_k\left(\frac{1}{2}\right) = \frac{1}{2} \cdot \left(1 - Tie\left(k, \frac{1}{2}\right)\right) - \frac{1}{2} < \frac{1}{2} - \frac{1}{2} = 0, \quad \text{and}$$

$$\lim_{k\to\infty} h_k\left(\frac{1}{2}\right) = \lim_{k\to\infty}\left(\frac{1}{2} \cdot \left(1 - Tie\left(k, \frac{1}{2}\right)\right) - \frac{1}{2}\right) = \left(\frac{1}{2} - 0\right) - \frac{1}{2} = 0.$$

$\square$

Claims 2-3 imply that for each $k \geq 2$ the equation $h_k(p) = 0$ has a unique solution in the interval $(0,1)$, that this solution $p(k)$ corresponds to an $S(k)$ globally asymptotically stable state, that it satisfies $p(k) < 0.5$, and that it is increasing in k.

To complete the proof of Theorem 1, it remains to show that $p(2) > 0.28$. This inequality is an immediate corollary of the fact that for $p = 0.28$

$$h_2(p = 0.28) = 2p(1-p)^3 + p^2(1-p^2) - p \approx 0.001 > 0. \qquad \square$$

Figure 1 shows the $S(k)$ payoff sampling dynamics for various values of k.

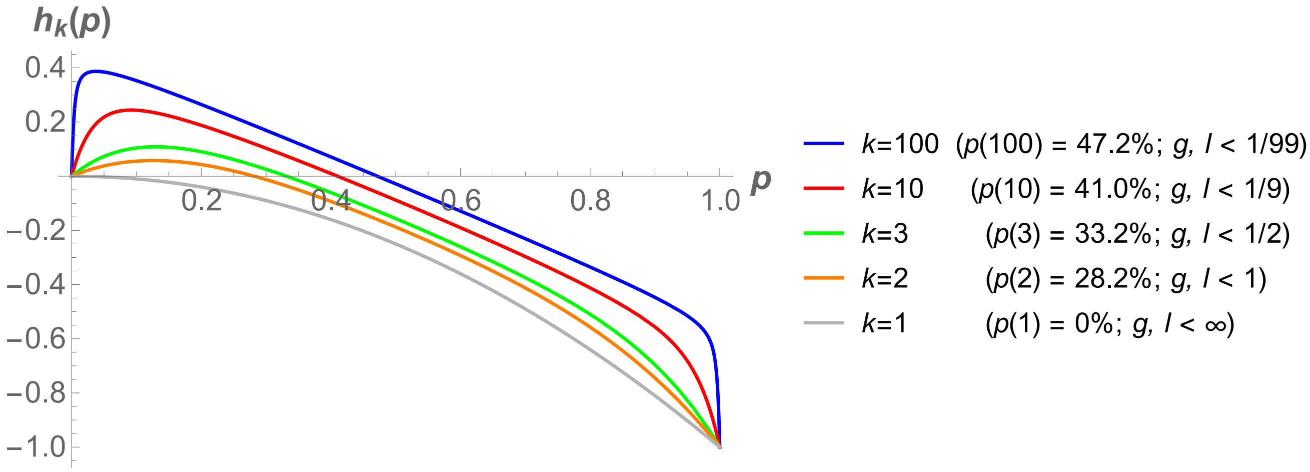

Figure 1: The function h_k and its zero $p(k)$ for various values of k.

Remark 1. For fixed $g, l > 0$, the condition $g, l < \frac{1}{k-1}$ does not hold for large k. In particular, in the limit $k \to \infty$ defection becomes globally asymptotically stable (see Sandholm et al., 2020, Prop. 4.3). The main motivation of the current paper is the study of small k's, which are plausibly the ones used by agents in real-life interactions, and for which the condition $g, l < \frac{1}{k-1}$ is, arguably, often satisfied.

3.3 Stability of Defection

Next we characterize the conditions under which defection is an $S(k)$ asymptotically stable equilibrium. This characterization supplements the analysis in the previous subsection in that it holds also for large values of g and l.

Proposition 2. *Assume that* $(k-1)l \neq 1$.[2] *Everyone defecting is an asymptotically stable state* $S(k)$ *equilibrium iff* $k = 1$ *or* $l > \frac{1}{k-1}$.

The proposition is implied by Theorem 2 below (and by the results of Sandholm et al., 2020). For completeness, we provide a direct sketch of proof.

Sketch of proof. The case of $k = 1$ is implied by Proposition 1. We are left with the case $k > 1$. Consider a population state in which $\epsilon << 1$ of the agents cooperate, and the

[2]The stability of defection in the borderline case of $l = \frac{1}{k-1}$ depends on the tie-breaking rule, because observing a single c in the c-sample and no c's in the d-sample gives a tie between the two samples. If one assumes a uniform tie-breaking rule, then action c wins with a probability $\frac{k}{2}\epsilon - O(\epsilon^2)$, which is larger than ϵ iff $k > 2$. Thus, with this rule, defection is stable if $k = 2$ and unstable if $k > 2$.

remaining $1-\epsilon$ agents defect. A revising agent usually sees all the opponents defecting both in the c-sample and in the d-sample. With a probability of approximately $k\epsilon$, the agent sees a single cooperation in the c-sample and no cooperation in the d-sample, so c yields a mean payoff of $(1-(k-1)l)/k$ and d yields 0. The former is higher iff $l < \frac{1}{k-1}$. Thus, if the last inequality holds, then the prevalence of cooperation gradually increases, and the population moves away from the state where everyone defects. By contrast, if $l > \frac{1}{k-1}$, then cooperation yields the higher mean payoff only if the c-sample includes at least two cooperators, which happens with a negligible probability of order $O(\epsilon^2)$. Therefore, in this case, cooperation gradually dies out, and the population converges to the state where everyone defects. □

4. Stability of Strict Equilibria

In this section we extend Proposition 2 by presenting a necessary and sufficient condition for asymptotic stability of any pure profile in any generic symmetric game.

4.1 Definitions

A symmetric game with payoff function $u : A \times A \to \mathbb{R}$ is *generic* if for any two finite sequences of action profiles $\left(a_j^1, a_j^2\right)_j$ and $\left(\tilde{a}_j^1, \tilde{a}_j^2\right)_j$ of equal length the equality

$$\sum_j u\left(a_j^1, a_j^2\right) = \sum_j u\left(\tilde{a}_j^1, \tilde{a}_j^2\right)$$

implies that each sequence is a permutation of the other. Observe that if each entry in the payoff matrix is independently drawn from a continuous (atomless) distribution, then the resulting random symmetric game is generic with probability one. Clearly, in a generic game, every pure Nash equilibrium is a strict equilibrium.

Definition 4. Let a^* be an equilibrium strategy in a generic two-player symmetric game. Let $a, a' \in A \setminus \{a^*\}$. Then:

1. Action *a directly $S(k)$ supports a' against a^** if

$$u(a', a) + (k-1) \cdot u(a', a^*) > k \cdot u(a^*, a^*).$$

2. Action a $S(k)$ *supports* a' *by spoiling* a^* if

$$k \cdot u(a', a^*) > (k-1) \cdot u(a^*, a^*) + u(a^*, a) \quad \text{and} \quad a' \in \operatorname*{argmax}_{b \neq a^*} u(b, a^*).$$

Action a $S(k)$ *supports* a' *against* a^* if at least one of the two conditions 1 and 2 holds, and $S(k)$ *single* or *double* supports it if only one or both of them, respectively, hold.

Action a directly supports a' (called being a supporter in Sandholm et al., 2020, p.12) if a single appearance of a in the a'-sample (with all other actions being a^*) is sufficient to make the mean payoff larger than that yielded by a^*, and thus to make it the largest payoff. Action a supports a' by spoiling a^* (called being a benefiting spoiler in Sandholm et al., 2020) if a single appearance of a in the a^*-sample is sufficient to make the mean payoff smaller than that yielded by a'. This makes the latter the largest mean payoff iff a' is the second-best reply to a^*. Note that, because the game is generic, the second-best reply is unique; the set $\operatorname{argmax}_{b \neq a^*} u(b, a^*)$ is actually a singleton.

Observe that $S(k)$ support is "easier" the smaller k is. That is, if action a $S(k)$ supports action a' against a^*, then it also $S(k')$ supports it for all $k' < k$.

4.2 Result

Next we characterize the condition for an action to be $S(k)$ asymptotically stable for all $k \geq 2$.

Theorem 2. *In a generic symmetric two-player game, and for any $k \geq 2$, everyone playing action a^* is $S(k)$ asymptotically stable $S(k)$ iff*

I. *(a^*, a^*) is a strict equilibrium, and*

II. *for every nonempty subset of actions $A' \subseteq A \backslash \{a^*\}$ there is an action $a' \in A'$ that is not $S(k)$ supported by any action in A'.*

The same is true with "not S(k) supported by" in condition 2 replaced with "does not S(k) support".

Sketch of proof. An action a^* that is not a symmetric Nash equilibrium cannot be an $S(k)$ equilibrium (because it is not a best-reply to itself, and therefore $w_{a^*,k}(e_{a^*}) = 0$, which contradicts e_{a^*} being a rest point). Thus, we can assume that a^* is a symmetric Nash equilibrium (actually, strict equilibrium, as the game is generic).

Suppose that there is a nonempty subset of actions $A' \subseteq A \backslash \{a^*\}$, with cardinality $n \geq 1$, such that each action $a' \in A'$ is supported by some action in A'. Consider an initial population state α in which a fraction $1-\epsilon$ of the agents play a^* and $\frac{\epsilon}{n}$ play each of the actions in A'. Since each $a' \in A'$ is supported by some action in A', it has a probability of approximately $k\frac{\epsilon}{n}$ of having that action appearing in the sample and thus making the mean payoff yielded by a' the highest one. It follows that $w_{a',k}(\alpha) = k\frac{\epsilon}{n} > \frac{\epsilon}{n} = \alpha_{a'}$ for all $a' \in A'$. Thus the frequency of all actions in A' increases. The formal proof (Appendix A.4) formalizes this intuition, by studying the Jacobian matrix at at e_{a^*}, and showing that it admits an eigenvalue larger than 1.

Next, suppose the converse, that any nonempty subset $A' \subseteq A \backslash \{a^*\}$ includes an action a' that is not $S(k)$ supported by any action in A'. Consider a state in which $1-\epsilon$ of the agents play a^*. We know that there exists action a' that is not $S(k)$ supported by any action in $A \backslash \{a^*\}$. This implies that the probability of action a' having the maximal mean payoff in an agent's sample is $O(\epsilon^2)$, and, thus, $w_{a',k}(\alpha) = O(\epsilon^2)$. As the frequency of a' becomes negligible, we can iterate the argument for $A' = A \backslash \{a^*, a'\}$, and finding another action a'', for which $w_{a'',k}(\alpha) = O(\epsilon^2)$, etc. The formal proof (Appendix A.4) shows that (a) Condition (II) implies that all the eigenvalues of the Jacobian matrix are negative, and (b) the phrase "not S(k) supported by" can be replaced with "does not S(k) support". □

We conclude this subsection with a few observations on Theorem 2:

1. It is easy to verify that conditions (I-II) imply asymptotic stability also for $k = 1$. The converse is not true, as is demonstrated by defection in the Prisoner's Dilemma.

2. Theorem 2 can be extended to symmetric n-player games (with minor adaptations to the proof). In this extension, the condition $k \geq 2$ is replaced by $(n-1) \cdot k \geq 2$.

3. Theorem 2 can be extended to non-generic symmetric games, by slightly weakening the "only if" conditions to be stated as (I) (a^*, a^*) is a Nash equilibrium, and (II) $\exists a' \in A'$ that is not $S(k)$ *weakly* supported by any action in A', where weak support is defined by replacing the strict inequalities in Definition 4 by weak inequalities.

4.3 Comparison with Sandholm et al. (2020)

We conclude by comparing Theorem 2 with the conditions for stability presented in Sandholm et al. (2020, Section 5) (which, in turn, strengthen the conditions presented in Sethi, 2000).[3]

The first result presents sufficient conditions for instability.

[3] Our comparison focuses on revising agents who test all feasible actions. The more general setup in Sandholm et al. allows dynamics in which revising agents test a subset of actions.

Adaptation of Proposition 5.4 (Sandholm et al., 2020) Let (a^*, a^*) be a strict equilibrium in a symmetric game and $k \geq 2$. Then a^* is not $S(k)$ asymptotically stable if either:

1. $\exists A' \subseteq A \backslash \{a^*\}$ such that every $a' \in A'$ is directly supported by some action in A'; or

2. $\exists A' \subseteq A \backslash \{a^*\}$ such that every action $a' \in A'$ supports some action in A'.

Theorem 2 strengthens this result by removing 'directly' from condition 1, thus weakening the condition. Moreover, this weaker condition (call it 1′) is actually equivalent to condition 2. This follows immediately from Theorem 2, as according to that result both the negation of 1′ and that of 2 are necessary *and sufficient* conditions for asymptotic stability.

Sandholm et al. (2020) present the following sufficient condition for stability.

Definition 5. *Let $a, a' \in A \backslash \{a^*\}$.* Action *a tentatively $S(k)$ supports a' by spoiling a^** if

$$k \cdot u(a', a^*) > (k-1) \cdot u(a^*, a^*) + u(a^*, a).$$

Adaptation of Prop. 5.9 (Sandholm et al., 2020) Let (a^*, a^*) be a strict equilibrium in a generic symmetric game and let $k \geq 2$. Then a^* is $S(k)$ asymptotically stable if

3. there exists an ordering of $A \backslash \{a^*\}$ such that no action $a \neq a^*$ either directly $S(k)$ supports or tentatively $S(k)$ supports by spoiling any weakly-higher action $a' \neq a^*$.

Theorem 2 strengthens this result by allowing removing 'tentatively' from condition 3, thus weakening the condition and making it *necessary* and sufficient for stability.[4] Sufficiency still holds because the weaker condition (call it 3′) implies that the lowest action in every subset $A' \subseteq A \backslash \{a^*\}$ does not $S(k)$ support any action in A'. Necessity holds because it is not very difficult to see that 3′ is implied by condition 2 in Theorem 2. (Recursively, remove from $A \backslash \{a^*\}$ an element that is not $S(k)$ supported by any of the remaining elements.)

A. Appendix

A.1 Proof of Claim 1

Recall that $p \equiv \alpha_c$ and $1-p \equiv \alpha_d$ denote the proportion of agents in the population playing action c and d, respectively. An agent's c-sample includes k actions of the opponents.

[4] Condition 3 itself is not necessary for asymptotic stability. In the game defined by the following payoff matrix, action a^* is $S(2)$ asymptotically stable as it satisfies the condition in Theorem 2, yet it does not satisfy condition 3 due to action a'' tentatively supporting itself by spoiling.

a^*	8	9	3
a'	7	5	2
a''	6	4	1

If $j \in \{0,..,k\}$ of these are c, then the agent's mean payoff (when playing c) is $j-(k-j)l$. Similarly, if $j' \in \{0,..,k\}$ of the sampled actions in the agent's d-sample are c, then the mean payoff (when playing action d) is $j'(1+g)$. The difference between the two payoffs is

$$(j-(k-j)l)-j'(1+g)=j-j'-((k-j)l+j'g).$$

This expression is clearly negative if $j \le j'$. If $j \ge j'+1$, it is positive, since $(k-j)l+j'g \le \frac{k-j+j'}{k-1} \le 1$ and the first inequality holds as equality only if $j=k$ and $j'=0$ while the second one does so only if $j=j'+1$.

These conclusions prove that the c-sample yields a superior payoff iff it includes more cooperations than the d-sample does. As the number of cooperators in each sample has a binomial distribution with parameters k and p, we conclude that $w_{c,k}=Win(k,p)$.

A.2 Proof of Claim 2

A binomial random variable with parameters k,p has a degenerate distribution if $p \in \{0,1\}$, so $Win(k,0)=Win(k,1)=0$. From Eq. (3.2) we get $h_k(0)=0-0=0$, $h_k(1)=0-1=-1$. To show that $h'_k(0)>0$ for $k=2,3,4,\dots$, we use the fact that

$$\begin{aligned}
h_k(p) &= Win(k,p)-p = 0.5(1-Tie(k,p))-p = 0.5(1-\sum_{j=0}^{k}(f_{k,p}(j))^2)-p) \\
&= 0.5(1-((1-p)^{2k}+O(p^2))-p \qquad (O(p^2) \text{ denotes the terms with degree} \ge 2) \\
&= 0.5(1-(1-2pk+O(p^2))-p+O(p^2)=kp-p+O(p^2)=(k-1)p+O(p^2)
\end{aligned}$$

From the above expression, it follows that $h'_k(0)=k-1>0$ for $k>1$.

A.3 Proof of Claim 3

Let $\{A_{j,p}\}_{j=1}^k$, $\{B_{j,p}\}_{j=1}^k$ be $2k$ independent Bernoulli random variables with parameter p. Then $X_{k,p}=\sum_{i=1}^k A_{j,p}$ and $Y_{k,p}=\sum_{j=1}^k B_{j,p}$ are i.i.d. binomial random variables with parameters k and p. Eq. (3.2) can be expressed as

$$\begin{aligned}
h_k(p) &= P(X_{k,p}>Y_{k,p})-p \\
&= \frac{1}{2}\Big(P(X_{k,p}>Y_{k,p})+P(X_{k,p}<Y_{k,p})\Big)-p \quad (\text{since } X_{k,p} \text{ and } Y_{k,p} \text{ are } i.i.d.) \\
&= \frac{1}{2}\Big(1-P(X_{k,p}=Y_{k,p})\Big)-p. \qquad \text{(A.1)}
\end{aligned}$$

For $j=1,2,\dots,k$, let $Z_{j,p}=A_{j,p}-B_{j,p}$. Clearly, $\{Z_{j,p}\}_{j=1}^n$ are i.i.d., with distribution given by

$$P(Z_{j,p}=-1)=P(Z_{j,p}=1)=pq \;\text{ and }\; P(Z_{j,p}=0)=1-2pq,$$

where $q = 1 - p$.

Consider the characteristic function $\varphi(t;p)$ of $Z_p^k = X_{k,p} - Y_{k,p} = \sum_{j=1}^{k} Z_{j,p}$:

$$\begin{aligned}
\varphi_k(t;p) &= \mathbb{E}\left[e^{itZ_p^k}\right] = \mathbb{E}\left[e^{it\sum_{j=1}^{k} Z_{j,p}}\right] = \left(\mathbb{E}[e^{itZ_{1,p}}]\right)^k \\
&= \left(e^{it(-1)}pq + e^{it(1)}pq + e^{it(0)}(1-2pq)\right)^k = \left(1 + pq(e^{-it} + e^{it} - 2)\right)^k \\
&= (1 + 2pq(\cos(t) - 1))^k \qquad (\text{since } e^{-it} + e^{it} = 2\cos(t)) \\
&= \left(1 - 4p(1-p)\sin^2\left(\frac{t}{2}\right)\right)^k \qquad \left(\text{since } q = 1 - p \text{ and } \cos(t) = 1 - 2\sin^2\left(\frac{t}{2}\right)\right)
\end{aligned}$$

The base of the last exponent, with power k, is an expression that is convex as a function of p and lies between 0 and 1, strictly so if $0 < p < 1$ and t is not a whole multiple of π. Therefore, the same is true for $\varphi_k(t;p)$ and, if $0 < p < 1$ and t is not a whole multiple of π, $\varphi_k(t;p) > \varphi_{k+1}(t;p)$ and $\lim_{k\to\infty} \varphi_k(t;p) = 0$. It follows, since (see Fact 1)

$$P(X_{k,p} = Y_{k,p}) = P(Z_p^k = 0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \varphi_k(t;p)dt$$

and in view of Eq. (A.1), that $h_k(p)$ is concave and, for $0 < p < 1$, the sequence $(h_k(p))_{k=1}^{\infty}$ is strictly increasing and tends to $\frac{1}{2} - p$. This completes the proof.

Fact 1. $P(Z_p^k = 0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \varphi_k(t;p)dt$

Proof of Fact 1. From the definition of $\varphi_k(t;p)$, we have the following:

$$\begin{aligned}
\int_{-\pi}^{\pi} \varphi_k(t;p)dt &= \int_{-\pi}^{\pi} \mathbb{E}\left[e^{itZ_p^k}\right]dt = \mathbb{E}\left[\int_{-\pi}^{\pi} e^{itZ_p^k}dt\right] = \mathbb{E}\left[\int_{-\pi}^{\pi} e^{itZ_p^k}\left(\mathbb{1}_{\{Z_p^k=0\}} + \mathbb{1}_{\{Z_p^k\neq 0\}}\right)dt\right] \\
&= \mathbb{E}\left[\int_{-\pi}^{\pi} e^{itZ_p^k}\mathbb{1}_{\{Z_p^k=0\}}dt\right] + \mathbb{E}\left[\int_{-\pi}^{\pi} e^{itZ_p^k}\mathbb{1}_{\{Z_p^k\neq 0\}}dt\right] \\
&= \mathbb{E}\left[\int_{-\pi}^{\pi} 1\cdot\mathbb{1}_{\{Z_p^k=0\}}dt\right] + \mathbb{E}[0] = \mathbb{E}\left[2\pi\cdot\mathbb{1}_{\{Z_p^k=0\}}\right] = 2\pi\cdot P(Z_p^k = 0)
\end{aligned}$$

From the above series of equalities, it follows that $P(Z_p^k = 0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \varphi_k(t;p)dt$.

A.4 Proof of Theorem 2

For completeness, we present all details of the proof, although various steps are analogous to arguments presented in the proofs of (Sandholm et al., 2020, Section 5).

For the set $B = A\backslash\{a^*\}$, let T be the non-negative $|B| \times |B|$ matrix whose element in row

a' and column a is

$$T_{a'a} = \begin{cases} 2 \text{ if action } a \text{ double } S(k) \text{ supports action } a' \\ 1 \text{ if action } a \text{ single } S(k) \text{ supports action } a' \\ 0 \text{ if action } a \text{ does not } S(k) \text{ support action } a' \end{cases}$$

We will now compute the Jacobian of the k-payoff sampling dynamic (2.2) at the pure strict Nash equilibrium e_{a^*}. Let ϵ be a small positive number, and let the frequency α_{a^*} of the strict equilibrium a^* in the population be $1 - \epsilon$. Denote by $\beta \equiv \alpha|_B$ the frequencies of the actions in set B, that is, $\beta_{a'} = \alpha_{a'}$ for all $a' \in B$. Clearly, $\sum_{a' \in B} \beta_{a'} = \epsilon$. This implies that the Euclidean norm of β is of the order ϵ, i.e., $|\beta| = \left(\sum_{a' \in B} \beta_{a'}^2\right)^{\frac{1}{2}} = O(\epsilon)$.

When $a' \in B$ is tested k times at the population state α, with a very high probability (namely, $(1-\epsilon)^k$) it encounters the equilibrium action a^* each time. The probability that the equilibrium action a^* is encountered $k-2$ or fewer times is of the order ϵ^2 or higher. All these higher order terms can be neglected for stability analysis. The probability that the non-equilibrium action $a' \in B$ yields the best payoff is roughly equal to the probability that it yields a better payoff than the equilibrium action a^* does when both actions are tested k times. For a' to beat a^*, at least one of the two following cases should occur.

Case 1: When a' is tested, one of the k opponents plays a non-equilibrium action a and the remaining opponents play a^*. Action a' obtains the maximal mean payoff if the first condition in Definition 4 is satisfied, that is, $u(a', a) + (k-1) \cdot u(a', a^*) > k \cdot u(a^*, a^*)$.

Case 2: When a^* is tested, one of the k opponents plays a' and the others play a^*. Action a' obtains the maximal mean payoff if the second condition in Definition 4 is satisfied, that is, $k \cdot u(a', a^*) > (k-1) \cdot u(a^*, a^*) + u(a^*, a)$ and $a' \in \operatorname{argmax}_{b \neq a^*} u(b, a^*)$.

Denoting the higher order terms (ϵ^2 and higher) by $O_{a'}(\epsilon^2)$, the probability that action a' yields the best payoff is then given by the following expression.

$$w_{a',k}(\alpha) = k \sum_{a \in B} T_{a'a} \alpha_a + O(\epsilon^2) = k \sum_{a \in B} T_{a'a} \beta_a + O_{a'}(|\beta|^2).$$

For $a' \in B$, the k-payoff sampling dynamic (2.2) can therefore be written as follows:

$$\dot{\beta}_{a'} = w_{a',k}(\alpha) - \beta_{a'} = k \sum_{a \in B} T_{a'a} \beta_a + O_{a'}(|\beta|^2) - \beta_{a'},$$

or in matrix notation,

$$\dot{\beta} = f(\beta) \equiv (kT - I)\beta + O(|\beta|^2), \tag{A.2}$$

where I is a $|B| \times |B|$ identity matrix and $O(|\beta|^2)$ is a $|B|$-dimensional vector with elements $O_{a'}(|\beta|^2)$ (i.e., all the elements are of the order $|\beta|^2$ or higher). Let J denote the Jacobian matrix evaluated at the origin. By definition, we have:

$$J = \frac{\partial f(\beta^*)}{\partial \beta}\bigg|_{\beta^* = \underbrace{(0, 0, \ldots, 0)}_{|B|-1 \text{ zeros}}} = kT - I.$$

The asymptotic stability of the system (A.2) can now be analyzed by examining the eigenvalues of the Jacobian matrix J.

A sufficient condition for a^* to be an $S(k)$ asymptotically stable state is that all the eigenvalues of J have negative real part (see, e.g., Sandholm, 2010, Corollary 8.C.2). A sufficient condition for it *not* to be an $S(k)$ asymptotically stable state is that at least one of the eigenvalues has positive real part. The first sufficient condition holds, in particular, if the only eigenvalue of T is zero, in other words, if the spectral radius ρ of that matrix is 0, as this condition means that the only eigenvalue of J is -1. The second sufficient condition holds if $\rho \geq 1$. This is because the spectral radius of a nonnegative matrix is an eigenvalue with a nonnegative eigenvector (Johnson and Horn, 1985, Theorem 8.3.1), so $\rho \geq 1$ implies that J has the eigenvalue $k\rho - 1 \geq 2 \cdot 1 - 1 > 0$, with a corresponding nonnegative eigenvector. It therefore suffices to show that $\rho = 0$ holds if condition 2 in the theorem holds, and $\rho \geq 1$ holds if that condition does not hold. Observe that condition 2 (resp., the alternative phrasing of condition 2 with "does not S(k) support") can be rephrased as follows: every principal submatrix of T has a row (resp., column) where all entries are zero. Therefore, to complete the proof of the theorem, it remains only to establish the following.

Fact 2. *Let M be a square matrix of nonnegative integers, and ρ its spectral radius. If every principal submatrix of M has a row of zeros, then $\rho = 0$. Otherwise, $\rho \geq 1$. The same is true with 'row' replaced by 'column'.*

Proof of Fact 2. Suppose that $\rho > 0$. Let v be a corresponding nonnegative right eigenvector. Since $(Mv)_i = \rho v_i > 0$ for every index i with $v_i > 0$, the set α of all such indices defines a principal submatrix M' (obtained from M by deleting all rows and all columns with indices not in α) with the property that every row includes at least one nonzero entry.

Conversely, suppose that M has a principal submatrix M', defined by some set of indices α, with the above property. Let v be a column vector of 0's and 1's where an entry is 1 iff its index lies in α. It is easy to see that $Mv \geq v$. This vector inequality implies that $\rho \geq 1$ (Johnson and Horn, 1985, Theorem 8.3.2).

The proof for 'column' is obtained by replacing M with its transpose.

References

Burton-Chellew, M. N., El Mouden, C., and West, S. A. (2017). Social learning and the demise of costly cooperation in humans. *Proceedings of the Royal Society B: Biological Sciences*, 284(1853):20170067.

Cárdenas, J., Mantilla, C., and Sethi, R. (2015). Stable sampling equilibrium in common pool resource games. *games*, 6(3):299–317.

Chmura, T. and Güth, W. (2011). The minority of three-game: An experimental and theoretical analysis. *Games*, 2(3):333–354.

Heller, Y. and Mohlin, E. (2018). Social learning and the shadow of the past. *Journal of Economic Theory*, 177:426–460.

Johnson, C. R. and Horn, R. A. (1985). *Matrix analysis*. Cambridge University Press Cambridge.

Kosfeld, M., Droste, E., and Voorneveld, M. (2002). A myopic adjustment process leading to best-reply matching. *Games and Economic Behavior*, 40(2):270–298.

Kreindler, G. E. and Young, H. P. (2013). Fast convergence in evolutionary equilibrium selection. *Games and Economic Behavior*, 80:39–67.

Mantilla, C., Sethi, R., and Cárdenas, J. C. (2018). Efficiency and stability of sampling equilibrium in public goods games. *Journal of Public Economic Theory*.

Mengel, F. (2018). Risk and temptation: A meta-study on prisoner's dilemma games. *The Economic Journal*, 128(616):3182–3209.

Mikekisz, J. and Ramsza, M. (2013). Sampling dynamics of a symmetric ultimatum game. *Dynamic Games and Applications*, 3(3):374–386.

Nax, H. H., Burton-Chellew, M. N., West, S. A., and Young, H. P. (2016). Learning in a black box. *Journal of Economic Behavior & Organization*, 127:1–15.

Nax, H. H. and Perc, M. (2015). Directional learning and the provisioning of public goods. *Scientific reports*, 5(1):1–6.

Osborne, M. J. and Rubinstein, A. (1998). Games with procedurally rational players. *American Economic Review*, pages 834–847.

Oyama, D., Sandholm, W. H., and Tercieux, O. (2015). Sampling best response dynamics and deterministic equilibrium selection. *Theoretical Economics*, 10(1):243–281.

Ramsza, M. (2005). Stability of pure strategy sampling equilibria. *International Journal of Game Theory*, 33(4):515–521.

Rowthorn, R. and Sethi, R. (2008). Procedural rationality and equilibrium trust. *The Economic Journal*, 118(530):889–905.

Salant, Y. and Cherry, J. (2019). Statistical inference in games. Technical report, mimeo.

Sandholm, W. H. (2001). Almost global convergence to p-dominant equilibrium. *International Journal of Game Theory*, 30(1):107–116.

Sandholm, W. H. (2010). *Population games and evolutionary dynamics*. MIT press.

Sandholm, W. H., Izquierdo, S. S., and Izquierdo, L. R. (2019). Best experienced payoff dynamics and cooperation in the centipede game. *Theoretical Economics*, 14(4):1347–1385.

Sandholm, W. H., Izquierdo, S. S., and Izquierdo, L. R. (2020). Stability for best experienced payoff dynamics. *Journal of Economic Theory*, 185:104957.

Sethi, R. (2000). Stability of equilibria in games with procedurally rational players. *Games and Economic Behavior*, 32(1):85–104.

Sethi, R. (2019). Procedural rationality in repeated games. Unpublished manuscript, Barnard College, Columbia University.

Spiegler, R. (2006a). Competition over agents with boundedly rational expectations. *Theoretical Economics*, 1(2):207–231.

Spiegler, R. (2006b). The market for quacks. *The Review of Economic Studies*, 73(4):1113–1131.