THE PINCHAS SAPIR CENTER
FOR DEVELOPMENT
Tel Aviv University

המרכז לפיתוח ע"ש פנחס ספיר
ליד אוניברסיטת תל אביב
"עמותה רשומה"
580010221

THE PINHAS SAPIR CENTER FOR DEVELOPMENT
TEL AVIV UNIVERSITY

Sampling-Based Reasoning and the Emergence of Reciprocity

Ran Spiegler

Discussion Paper No. 2-2022

Sampling-Based Reasoning and the Emergence of Reciprocity[*]

Ran Spiegler[†]

December 11, 2021

Abstract

I study an infinite-horizon trust game, in which at each period, a distinct player chooses whether to put trust in the next player. Players are limited to bounded-recall strategies. Each player forms his belief regarding his opponent's strategy on the basis of sample data, drawn from the long-run play path. In equilibrium, players best-reply to their belief. I demonstrate how the combination of sampling error and the representative-sample aspect of players' sampling procedure lead to the emergence of reciprocal behavior.

[*] Financial support from the Sapir Center is gratefully acknowledged. I thank Nathan Hancart for helpful feedback.

[†] Tel Aviv University, University College London and CFM. URL: http://www.tau.ac.il/~rani. E-mail: rani@tauex.tau.ac.il.

1 Introduction

How to sustain long-run trust in dynamic interactions is one of the most thoroughly studied questions in Game Theory. This literature has emphasized the question of whether cooperative behavior is consistent with Nash equilibrium and its refinements (see Mailath and Samuelson (2006) for a textbook treatment). A smaller literature addressed the question of how players may *learn* to cooperate through some dynamic, non-equilibrium learning process (e.g. Kalai and Lehrer (1993)).

This paper tackles the problem of learning and sustaining cooperative behavior from the perspective of a smaller literature, which seeks to fuse the learning and equilibrium perspectives (Osborne and Rubinstein (1998), Spiegler (2005), Salant and Cherry (2020), Goncalves (2020)). According to this approach, learning - in the sense of drawing inferences from partial data - is an integral part of the definition of equilibrium behavior. Players extrapolate beliefs from sample data. Equilibrium behavior is the outcome of players best-replying to the beliefs extrapolated from their sample. In turn, the sample is drawn from the equilibrium data-generating process.

I apply this approach to a discrete-time, infinite-horizon trust game with sequential moves. At every period t, a distinct player (also called t) acts. His payoff depends only on his own action and the action of player $t+1$, according to a standard Prisoner's Dilemma payoff function. Thus, each player's dilemma is whether to put trust in the subsequent player. I restrict attention to finite-recall strategies: players can only condition their action on the m most recent actions. Under this restriction, there is a unique Nash equilibrium, in which players always defect.

To incorporate learning into a definition of equilibrium, think of a scenario in which the game has been played for many periods. Players lack access to the entire history. Instead, they obtain data about a random chunk of the historical play path. They regard this chunk as a sample that enables them to learn players' strategy in this game - i.e., how players condition their action on the m-truncated history that precedes them. For instance, suppose that $m = 1$ and a player extracts the following action sequence

110110000000010111110 from the long-run play path (the digits 0 and 1 indicate defection and cooperation, respectively). A natural inference from this 21-period chunk is that players are cooperative with probability 60% following cooperative play (because in this chunk, 1 is played 10 times, followed by 1 in 6 out of these 10 cases), and that players are cooperative with probability 30% following defection (because in this chunk, 0 is played 10 times - not including the last observation - followed by 1 in 3 out of these 10 cases).

Literally formalizing this scenario is complex. Instead, I develop a more tractable modeling approximation. To motivate it, suppose all players follow the same mixed strategy (namely, a function that assigns a probability of cooperation to every m-truncated history). Assuming the strategy has full support, it induced a unique invariant distribution over m-truncated histories. Imagine that each player obtains access to a sample of n independent observations of how players respond to m-truncated histories. The sample is *representative* - that is, each truncated history is sampled in proportion to its long-run frequency under the invariant distribution. The representative-sample assumption approximates the idea that if players have access to an arbitrary chunk of the long-run play path, the are more likely to encounter a particular truncated history if it has a high invariant probability. For each truncated history, the player observes a normally distributed variable. Its moments are defined by the sample average of independent draws from the action mixture that the equilibrium strategy assigns to the truncated history in question. This is a normal approximation of the idea that for every sample point, the player observes an independent draw from the Bernoulli distribution induced by the equilibrium strategy at a given truncated history. The role of this smooth approximation is to avoid the problem that an exact representative sample will typically involve a fractional number of observations. The realization of this normal variable is the player's point forecast of the subsequent player's propensity to cooperate at the given history.

Thus, my modeling approximation involves two tricks. First, I assume that players obtain a representative sample of players' response to truncated histories, rather than a random chunk of the play path. Second, I use a Gaussian approximation of the otherwise discrete distribution of the average of independent draws from a Bernoulli distribution. The equilibrium condition is that players best-reply to their point forecasts, and that the random behavior that this sampling-based procedure generates coincides with the equilibrium strategy.

At present, I have partial characterization results for $m \leq 2$. The main result is that this equilibrium concept produces positive rates of cooperative behavior, such that players' propensity to cooperate is higher following cooperative play. This reciprocity effect arises because of the representative-sample assumption. The intuition is that when a pattern of behavior is more common, it accordingly receives large representation in players' sample. This makes players' assessment of how their opponent will react to this pattern more precise, having a narrower tail. This has implications for the probability of a high evaluation of the probability of cooperation following different patterns. Numerical simulations for $m = 3, 4, 5$ confirm this result, but establishing it analytically for general m remains an open problem.

Sampling-based reasoning thus makes reciprocity self-sustaining. Reciprocity changes the relative frequency of patterns that exhibit cooperative and non-cooperative behavior, which in turn affects the probability of tail events in the sampling of players' behavior following these different patterns. Although I have identified this effect in a simple dynamic trust game, I believe this insight is relevant for more complex interactions, such as models of repeated oligopolistic interaction.

2 The Model

Consider the following discrete-time, infinite-horizon, sequential-move game. It will be helpful to imagine time as stretching to infinity in both directions - i.e., $t = \ldots - 2, 1, 0, 1, 2, \ldots.$ At every period t, a *distinct* agent, referred to as player t, chooses an action $a_t \in \{0, 1\}$.

Player t's payoff is purely a function of a_t and a_{t+1}, given by $u(a_t, a_{t+1}) = \frac{1}{c}a_{t+1} - a_t$, where $c < 1$ is a constant. This is a standard Prisoner's Dilemma payoff matrix: $a_t = 1$ means that player t decides to "put his trust" in player $t + 1$. This payoff function implies the following basic observation. If player t

believes that $a_{t+1} = 1$ with probability $p(a_t)$, then player t will weakly prefer to play $a = 1$ if and only if

$$\frac{1}{c} \cdot p(1) - 1 \geq \frac{1}{c} \cdot p(0) - 0$$

which is equivalent to

$$p(1) - p(0) \geq c$$

Players in this game have *limited recall.* They can only condition their action on m-truncated histories - i.e. the $m \geq 1$ most recent actions. The set of m-truncated histories is $H = \{0, 1\}^m$. For every m-truncated history $h = (a_{t-m}, ..., a_{t-1})$, (h, a) is a shorthand notation for the concatenated m-truncated history $(a_{t-m+1}, ..., a_{t-1}, a)$. A mixed strategy for any player t in this game is a function $f : H \to [0, 1]$, where $f(h)$ is the probability that $a_t = 1$ given the m-truncated history $h = (a_{t-m}, ..., a_{t-1})$.

Benchmark: Nash equilibrium
In the unique Nash equilibrium of this game, every player chooses $a = 0$ after every history. The reason is as follows. Fix a candidate Nash equilibrium. Define $m^* \leq m$ as the effective recall associated with this equilibrium - i.e. there is a player t who conditions his behavior on a_{t-m^*}, and there is no $m' > m^*$ for which this is the case. Suppose $m^* > 0$, and consider player t's reasoning. By the definition of m^*, this player knows that player $t + 1$ will not condition his behavior on a_{t-m^*}. Therefore, there is no reason for player t to condition his own behavior on a_{t-m^*}, because his own payoff only depends on a_t and a_{t+1}. This contradicts the definition of m^*. It follows that $m^* = 0$, which means that players never condition their behavior on the history. This makes $a = 0$ a best-reply for each player.

This finding relies on the limited-recall assumption. If players had perfect recall, cooperation could be sustained by a "grim" trigger strategy: play $a = 1$ if and only if all predecessors played $a = 1$.

Let us now present the sampling-based equilibrium concept for this game. Let $n > 0$ be an integer. Fix a strategy f. This strategy defines a discrete-time Markov process, in which the set of states is H, and the probabilities of transition from h into $(h,1)$ and $(h,0)$ are $f(h)$ and $1-f(h)$, respectively. If $f(h) \in (0,1)$ for every h, then the Markov process is irreducible and therefore has a unique invariant distribution over H, denoted α_f. Moreover, this distribution has full support.

Suppose that f induces a well-defined invariant distribution α_f. For every $h \in H$ and $a \in \{0,1\}$, define the following independently distributed, normal random variable:

$$\hat{f}(h,a) \sim N\left(f(h,a), \frac{f(h,a)(1-f(h,a))}{n\alpha_f(h,a)}\right) \tag{1}$$

This variable represents an individual player's estimate of the probability that the subsequent player will choose $a = 1$ following the m-truncated history (h,a).

The model has two free parameters: the payoff constant c, and the sample size n. We are now ready to define our equilibrium concept.

Definition 1 (Equilibrium) *Fix $\varepsilon \in (0,1)$. A strategy f is an ε-equilibrium if, for every $h \in H$,*

$$f(h) = \varepsilon \cdot \frac{1}{2} + (1-\varepsilon) \cdot \Pr(\hat{f}(h,1) - \hat{f}(h,0) \geq c)$$

where $\hat{f}$ is defined by (1). A strategy is an equilibrium if it is the limit of a sequence of ε-equilibria, where $\varepsilon \to 0$.

The procedure behind the equilibrium concept

Let us now describe the decision process that this definition approximates. When players' statistical behavior is consistent with the strategy f, this induces a play path in which the long-run frequency of m-truncated histories is given by α_f. This is due to the ergodicity property of the invariant distribution. Before taking an action, a player uses a sample from past play to form a belief regarding the subsequent player's behavior. Specifically, he may examine a chunk of the historical play path. This chunk is a random sample; yet on average, the frequencies of m-truncated histories in it will conform to

α_f. The player uses this chunk to generate a point estimate of the probability that players choose $a = 1$ as a function of the preceding m-truncated history. The equilibrium condition is that players' statistical behavior, given by f, is consistent with best-replying to their samples.

Literally modeling this process of random sampling of chunks of past play is complex. Therefore, for the sake of tractability, I make a number of modeling approximations. First, I envisage the player as if he has access to a *representative* sample of n truncated histories. The fractions of truncated histories in the sample are given by α_f. Second, I imagine that for each $a = 0, 1$, the player observes $n\alpha_f(h, a)$ independent draws from $f(h, a)$, and uses the average of these sample points as a point forecast of the probability that players choose $a = 1$ following (h, a).

Thus, one modeling simplification is to replace the image of observing a chunk of past play with the notion of a representative sample of truncated histories. The final modeling approximation replaces the discrete distribution of the sample average of $n\alpha_f(h, a)$ independent draws from a Bernoulli variable (having a success rate of $f(h, a)$) with a normal variable having the *same mean and variance.* This normal approximation is helpful because $n\alpha_f(h, a)$ is typically not a strictly positive integer, hence constructing an exactly representative sample is impossible.

Following this approximated process, player t acting after the m-truncated history h believes that $a_{t+1} = 1$ with probability $\hat{f}(h, a_t)$, and therefore he will choose $a = 1$ whenever $\hat{f}(h, 1) - \hat{f}(h, 0) \geq c$. The equilibrium condition is that the probability of this event coincides with $f(h)$. The definition of ε-equilibrium mixes the probability that the player best-replies to his forecast with the uniform distribution. Thus, the constant ε introduces an element of blind experimentation into the model, which ensures that the invariant distribution over H is unique and has full support.

Comment: The space of beliefs

The belief-formation procedure that players follow in this model carries an implicit assumption: each player t believes that the behavior of player $t + 1$ is measurable w.r.t the m-truncated history $(a_{t-m+1}, ..., a_t)$. This belief is correct because indeed, players' recall is bounded by m. However, if players did not incorporate this knowledge and relied purely on sample data, the existence of sampling error could lead players to infer a history dependence that exceeds the objective bound on players' recall. Therefore, the assumption is that players make use of their knowledge of the bound on recall when constructing the space of possible strategies in their estimation procedure.

3 Analysis for $m \leq 2$

This section is devoted to characterizing equilibrium in this model for $m \leq 2$. Let us begin with the simplest specification of this model.

The case of $m = 1$

When $m = 1$, $h \in \{0, 1\}$. Each player t believes that player $t + 1$ will not condition his action on a_{t-1}. Since player t's payoff only depends on a_t and a_{t+1}, it follows that player t's action will be history-independent. Therefore, in symmetric equilibrium, f is identified with a stationary probability of playing $a = 1$, denoted p. Therefore, the invariant distribution α_f assigns probability p to $h = 1$ and probability $1 - p$ to $h = 0$. It follows that

$$\begin{aligned} \hat{f}(1) &\sim N\left(p, \frac{p(1-p)}{np}\right) \\ \hat{f}(0) &\sim N\left(p, \frac{p(1-p)}{n(1-p)}\right) \end{aligned}$$

Note that

$$\frac{p(1-p)}{np} + \frac{p(1-p)}{n(1-p)} = \frac{1}{n}$$

Since $\hat{f}(1)$ and $\hat{f}(0)$ are independent normal variables with the same mean,

$$\hat{f}(1) - \hat{f}(0) \sim N(0, \frac{1}{n})$$

Therefore,

$$p = \Pr(\hat{f}(1) - \hat{f}(0) \geq c) = 1 - \Phi(c\sqrt{n})$$

We have thus pinned down the unique, stationary equilibrium. This result is intuitive. Sampling error can lead players to believe that playing $a = 1$ will increase the subsequent player's probability of playing $a = 1$. A decrease in c corresponds to a larger benefit from mutual trust, and therefore the propensity to play $a = 1$ increases. A rise in n results in a smaller sampling error, and therefore the potential for cooperative behavior diminishes.

Remark: Effective recall
As the case of $m = 1$ illustrates, the sequential-move structure of the game means that although players' recall is given by m, their effective recall is at most $m - 1$. That is, in equilibrium each player will only condition on the $m-1$ most recent actions. The reason is that player t's estimation procedure only considers strategies with recall m. Therefore, his procedure does not allow a_{t-m} to influence his forecast of the behavior of player $t + 1$. Since player t's payoff does not depend directly on a_{t-m}, this means that there is no scope for $f(h)$ to depend on the earliest action in h.

We can conclude that although f is nominally a function of the set of m-truncated histories H, the equilibrium concept means that we can restrict attention to functions f that are measurable w.r.t to the $m - 1$ most recent actions in each m-truncated history. For instance, as we saw, when $m = 1$, the equilibrium strategy is stationary (corresponding to no recall).

Remark: Reducing equilibrium to a system of equations
As we observed in our analysis of the $m = 1$ case, $\hat{f}(h, 1)$ and $\hat{f}(h, 0)$ are independent normal variables. Therefore, their difference $\hat{f}(h, 1) - \hat{f}(h, 0)$ is a normal variable $x(h)$, given by

$$x(h) \sim N\left(f(h,1) - f(h,0), \frac{f(h,1)(1 - f(h,1))}{n\alpha_f(h,1)} + \frac{f(h,0)(1 - f(h,0))}{n\alpha_f(h,0)}\right) \tag{2}$$

This means that an equilibrium f is a solution to the following system of equations:

$$f(h) = \Pr(x(h) \geq c) \tag{3}$$

where the cumulative distribution function of $\hat{f}(h, 1) - \hat{f}(h, 0)$ is determined by (2). The number of unknowns in this system is 2^{m-1} because, as we observed, $f(h)$ only depends on the $m-1$ most recent actions. (Note also that the object (h,a) does not record the earliest action in h.) The number of equations is also 2^{m-1}, because when h and h' have the same $m-1$ most recent realizations, $x(h)$ and $x(h')$ have the same distribution.

The following result tackles the case of $m=2$, and demonstrates that the equilibrium strategy in this case is not stationary. Indeed, it exhibits reciprocal behavior.

Proposition 1 *Let $m=2$. Then, in any equilibrium, $f(a_{t-2},a_{t-1})$ is strictly increasing in a_{t-1}.*

Proof. Recall that when $m=2$, f is effectively a function of the most recent action only. Accordingly, denote by f_a be the probability that $a_{t+1}=1$ conditional on $a_t=a$. In a similar vein, use the notation α_h for $\alpha_f(h)$. The system of equations (2)-(3) is reduced to

$$\begin{aligned} f_1 &= \Pr\left(\hat{f}(1,1)-\hat{f}(1,0)\geq c\right) \\ f_0 &= \Pr\left(\hat{f}(0,1)-\hat{f}(0,0)\geq c\right) \end{aligned}$$

where

$$\begin{aligned} \hat{f}(1,1)-\hat{f}(1,0) &\sim N\left(f_1-f_0,\frac{f_1(1-f_1)}{n\alpha_{11}}+\frac{f_0(1-f_0)}{n\alpha_{10}}\right) \\ \hat{f}(0,1)-\hat{f}(0,0) &\sim N\left(f_1-f_0,\frac{f_1(1-f_1)}{n\alpha_{01}}+\frac{f_0(1-f_0)}{n\alpha_{00}}\right) \end{aligned}$$

By the definition of α_f,

$$\begin{aligned} \alpha_{11} &= f_1\cdot(\alpha_{11}+\alpha_{01}) \\ \alpha_{10} &= (1-f_1)\cdot(\alpha_{11}+\alpha_{01}) \\ \alpha_{01} &= f_0\cdot(\alpha_{10}+\alpha_{00}) \\ \alpha_{00} &= (1-f_0)\cdot(\alpha_{10}+\alpha_{00}) \\ 1 &= \alpha_{00}+\alpha_{01}+\alpha_{10}+\alpha_{11} \end{aligned}$$

The solution for α_f is

$$\begin{aligned}\alpha_{11} &= \tfrac{f_1 f_0}{1+f_0-f_1} \\ \alpha_{10} &= \tfrac{f_0(1-f_1)}{1+f_0-f_1} \\ \alpha_{01} &= \tfrac{f_0(1-f_1)}{1+f_0-f_1} \\ \alpha_{00} &= \tfrac{(1-f_0)(1-f_1)}{1+f_0-f_1}\end{aligned}$$

Let us consider three cases. First, suppose $f_1 - f_0 = c$. Then, by the symmetry of the normal distribution, $f_1 = f_0 = \frac{1}{2}$, hence $f_1 - f_0 = 0 < c$, a contradiction.

Second, suppose $f_1 - f_0 > c$. Then, $f_1, f_0 > \frac{1}{2}$. Therefore, $\alpha_{11} > \alpha_{01}$ and $\alpha_{10} > \alpha_{00}$. It follows that $\hat{f}(1,1) - \hat{f}(1,0)$ and $\hat{f}(0,1) - \hat{f}(0,0)$ have the same mean and

$$Var(\hat{f}(0,1) - \hat{f}(0,0)) > Var(\hat{f}(1,1) - \hat{f}(1,0))$$

Since the mean lies above c,

$$f_1 = \Pr(\hat{f}(1,1) - \hat{f}(1,0) \geq c) > \Pr(\hat{f}(0,1) - \hat{f}(0,0) \geq c) = f_0$$

Finally, suppose $f_1 - f_0 < c$. Then, $f_1, f_0 < \frac{1}{2}$. Therefore, $\alpha_{11} < \alpha_{01}$ and $\alpha_{10} < \alpha_{00}$. It follows that $\hat{f}(1,1) - \hat{f}(1,0)$ and $\hat{f}(0,1) - \hat{f}(0,0)$ have the same mean, and

$$Var(\hat{f}(0,1) - \hat{f}(0,0)) < Var(\hat{f}(1,1) - \hat{f}(1,0))$$

Since the mean lies below c,

$$f_1 = \Pr(\hat{f}(1,1) - \hat{f}(1,0) \geq c) > \Pr(\hat{f}(0,1) - \hat{f}(0,0) \geq c) = f_0$$

This completes the proof. ■

The message of this result is that reciprocity emerges naturally when players form beliefs on the basis of sample data drawn from the equilibrium play path. For instance, suppose that cooperative play is less frequent than defecting. Then, the truncated history $h = 1$ is less frequent than the truncated history $h = 0$. A representative sample will therefore have fewer observations about how players act at the history $h = 1$ than at the history $h = 0$. As a result, the estimate of the benefit of playing $a = 1$ will have a fatter tail for the history $h = 1$ than for the history $h = 0$. The assumption that $a = 0$ is played more frequently than $a = 1$ means that c is above the mean of the estimates' distributions. As a result, a fatter tail means a higher probability of finding $a = 1$ to be optimal.

4 Two Variations

The discussion in the previous section traced the emergence of reciprocity in equilibrium to the representative-sample assumption - which in turn is a modeling approximation of the idea that players learn from a random slice of the equilibrium play path. In this section I present two variations on the model that cement this point.

4.1 A Final-Action Representative Sample

Suppose that when a player acts at the history h, he obtains a total of n observations, and allocates them into observations about what happens after the histories $(h, 1)$ and $(h, 0)$, with representative proportions. That is, he obtains $n \cdot f(h)$ independent draws from the Bernoulli distribution whose success rate is $f(h, 1)$, and $n \cdot (1 - f(h))$ independent draws from the Bernoulli distribution whose success rate is $f(h, 0)$. Our normal approximation of this description means that the only change in the basic model is that (2) is modified into

$$x(h) \sim N\left(f(h,1) - f(h,0), \frac{f(h,1)(1-f(h,1))}{nf(h)} + \frac{f(h,0)(1-f(h,0))}{n(1-f(h))}\right)$$

Let us guess a stationary equilibrium, such that $f(h) = b$ for every h.

Then, equilibrium is unique and given by:

$$
\begin{aligned}
x &\sim N\left(0, \frac{1}{n}\right) \\
b &= \Pr\left(x \geq c\right)
\end{aligned}
$$

This is the same stationary equilibrium we obtained in the case of $m = 1$ under the main model. The following result shows it is the only equilibrium when $m = 2$.

Proposition 2 *Let $m = 2$. Then, the stationary equilibrium is the unique equilibrium in the final-action-representative-sample case.*

Proof. Recall that when $m = 2$, f is effectively a function of the most recent action. Accordingly, let f_a be the probability that $a_{t+1} = 1$ conditional on $a_t = a$. Then,

$$
\begin{aligned}
f_1 &= \Pr\left(\hat{f}(1,1) - \hat{f}(1,0) \geq c\right) \\
f_0 &= \Pr\left(\hat{f}(0,1) - \hat{f}(0,0) \geq c\right)
\end{aligned}
$$

where

$$
\begin{aligned}
\hat{f}(1,1) - \hat{f}(1,0) &\sim N\left(f_1 - f_0, \frac{f_1(1-f_1)}{nf_1} + \frac{f_0(1-f_0)}{n(1-f_1)}\right) \\
\hat{f}(0,1) - \hat{f}(0,0) &\sim N\left(f_1 - f_0, \frac{f_1(1-f_1)}{nf_0} + \frac{f_0(1-f_0)}{n(1-f_0)}\right)
\end{aligned}
$$

Simplifying, we obtain

$$
\begin{aligned}
x(1) &= \hat{f}(1,1) - \hat{f}(1,0) \sim N\left(f_1 - f_0, \frac{1}{n}\left(1 - f_1 + \frac{f_0(1-f_0)}{(1-f_1)}\right)\right) \\
x(0) &= \hat{f}(0,1) - \hat{f}(0,0) \sim N\left(f_1 - f_0, \frac{1}{n}\left(\frac{f_1(1-f_1)}{f_0} + f_0\right)\right)
\end{aligned}
$$

Recall that

$$
f_a = \Pr(x(a) \geq c)
$$

Suppose $f_1 - f_0 > c$. Then, $f_1, f_0 > \frac{1}{2}$. Since $x(1)$ and $x(0)$ have the same mean which is above c, $f_1 > f_0$ only if the variance of $x(1)$ is lower than the variance of $x(0)$. Therefore,

$$1 - f_1 + \frac{f_0(1 - f_0)}{(1 - f_1)} < \frac{f_1(1 - f_1)}{f_0} + f_0$$

Rearranging, and using the assumption $f_1 > f_0$, we obtain

$$f_0 + f_1 < 1$$

contradicting our finding that $f_1, f_0 > \frac{1}{2}$.

Now suppose $f_1 - f_0 < 0$. Then, $f_1, f_0 < \frac{1}{2}$. Since $x(1)$ and $x(0)$ have the same mean which is below c, $f_1 > f_0$ only if the variance of $x(1)$ is larger than the variance of $x(0)$. Therefore,

$$1 - f_1 + \frac{f_0(1 - f_0)}{(1 - f_1)} > \frac{f_1(1 - f_1)}{f_0} + f_0$$

Rearranging, and using the assumption $f_1 < f_0$, we obtain

$$f_0 + f_1 > 1$$

contradicting our finding that $f_1, f_0 < \frac{1}{2}$.

Finally, suppose $0 < f_1 - f_0 < c$. Then, $f_1, f_0 < \frac{1}{2}$. Since $x(1)$ and $x(0)$ have the same mean which is above c, $f_1 > f_0$ only if the variance of $x(1)$ is larger than the variance of $x(0)$. Therefore,

$$1 - f_1 + \frac{f_0(1 - f_0)}{(1 - f_1)} > \frac{f_1(1 - f_1)}{f_0} + f_0$$

Since $f_0 < f_1 < \frac{1}{2}$, $f_1(1 - f_1) > f_0(1 - f_0)$. Therefore,

$$1 = 1 - f_1 + \frac{f_1(1 - f_1)}{(1 - f_1)} > 1 - f_1 + \frac{f_0(1 - f_0)}{(1 - f_1)} > \frac{f_1(1 - f_1)}{f_0} + f_0 > \frac{f_0(1 - f_0)}{f_0} + f_0 = 1$$

a contradiction. We have thus ruled out all possibilities of $f_1 \neq f_0$. ■

Thus, under the final-action-representative-sample model, reciprocal behavior cannot emerge when $m = 2$. Whether the same result holds for general m is an open question.

4.2 A Uniform Sample

Suppose that each player obtains an equal number of observations l for every history in H. That is, for every $h \in H$, the player observes l independent draws from the Bernoulli distribution whose success rate is $f(h)$. Our normal approximation of this description means that the only change in the basic model is that (1) changes into

$$\hat{f}(h,a) \sim N\left(f(h,a), \frac{f(h,a)(1-f(h,a))}{l}\right)$$

Therefore, (2) is modified into

$$x(h) \sim N\left(f(h,1) - f(h,0), \frac{f(h,1)(1-f(h,1))}{l} + \frac{f(h,0)(1-f(h,0))}{l}\right)$$

Guess a stationary ε-equilibrium, such that $f(h) = b$ for every h. Then, equilibrium is given by the following:

$$\begin{aligned} x &\sim N\left(0, \frac{2b(1-b)}{l}\right) \\ b &= \varepsilon \cdot \frac{1}{2} + (1-\varepsilon) \cdot \Pr(x \geq c) \end{aligned}$$

Let us see why there is some $b \in (0,1)$ that satisfies this pair of conditions. When $b = \frac{1}{2}$, the R.H.S of the second equation is below b because $c > 0$ and therefore $\Pr(x \geq c) < \frac{1}{2}$. When $b = 0$, the R.H.S is above b. Since both sides of the equation are continuous in b, the intermediate value theorem implies that there is a value of b that satisfies it. Note that in the $\varepsilon \to 0$, $b = 0$ is a solution - namely, no trust is displayed in equilibrium.

It follows that a stationary ε-equilibrium exists for every $\varepsilon > 0$, which means that reciprocity is not a necessary equilibrium phenomenon. Whether there exist multiple stationary equilibria and whether there exist non-stationary equilibria are open questions.

References

[1] Goncalves, Duarte (2020). "Sequential sampling and equilibrium". Working Paper.

[2] Kalai, Ehud, and Ehud Lehrer (1993). "Rational learning leads to Nash equilibrium." Econometrica 61, 1019-1045.

[3] Mailath, George J., and Larry Samuelson (2006). Repeated games and reputations: long-run relationships. Oxford university press.

[4] Osborne, Martin J., and Ariel Rubinstein (1998). "Games with procedurally rational players." American Economic Review 88, 834-847.

[5] Salant, Yuval, and Josh Cherry (2020). "Statistical inference in games." Econometrica 88, 1725-1752.

[6] Spiegler, Ran (2005). "Testing threats in repeated games." Journal of Economic Theory 121, 214-235.