Representative Sampling Equilibrium*

Tuval Danenberg[†] and Ran Spiegler[‡]

August 7, 2022

Abstract

We present an equilibrium concept based on the idea that agents evaluate actions using sample data drawn from the equilibrium distribution, where the number of observations about an alternative is proportional to its usage in a relevant population. Agents naively extrapolate from their data, using the sample mean payoff from each alternative as a predictor of their payoff from choosing it. The endogeneity of sample sizes gives rise to a novel equilibrium effect: Players' assessment of less frequently played actions is noisier. We study the implications of this effect in a single-agent, binary-choice model, as well as in various examples of games.

*Financial support from the Sapir Center is gratefully acknowledged. We thank Nathan Hancart and Yuval Salant for helpful feedback.

[†]Tel Aviv University. E-mail: tuvaldan@gmail.com

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

1 Introduction

Standard analysis of long-run behavior in strategic interactions follows the equilibrium approach. It focuses on steady-state behavior, with players acting as if they know long-run regularities. How players get to learn these regularities is left outside the scope of equilibrium analysis and relegated to separate models of learning in games (e.g., Fudenberg and Levine (1998)), which focus on players' dynamic responses to finitely many observations of past play.

A small literature — starting with Osborne and Rubinstein (1998), and including Spiegler (2006a,b), Salant and Cherry (2020) and Goncalves (2020) — has attempted to *fuse* these two approaches, by formulating equilibrium concepts in which learning from finite samples is intrinsic to equilibrium behavior. Players base their actions on some kind of inference from samples that are drawn from the equilibrium distribution, which in turn results from their own response to these samples. Equilibrium behavior is intrinsically random due to sampling errors.

Equilibrium concepts in this vein are based on fundamental assumptions regarding the procedure that players employ when forming their sample, and the procedure they use for drawing inferences from their sample. In particular, the choice of sampling procedure depends on the type of learning one wishes to capture: Are agents learning from passive observation or from active experimentation?

Osborne and Rubinstein (1998) and Salant and Cherry (2020) assumed uniform sampling, where each player samples each action K times (more precisely, she draws K independent sample points from each action's conditional outcome distribution). This uniform sampling procedure naturally fits situations in which players rely on deliberate experimentation to form beliefs. It is less appropriate as a description of situations in which players' equilibrium perceptions are based on passive, casual observational data. For example, consider a decision between two smartphone brands (in the absence of consumption externalities, this decision is non-strategic). To learn about the quality of each brand, an agent might ask some of her friends how happy they are with their current devices. In this case, her sample size for each brand will depend on its popularity among her friends.

To capture this kind of sampling-based decisions, we extend the Osborne-Rubinstein approach by assuming that each player's sample is not uniform but *representative*. In the simplest case of a single-agent choice problem, the decision maker constructs a sample of size n, such that the number of sample points about a given action a is $n \cdot q(a)$, where $q(a)$ is the frequency with which the action is taken in the population. Thus, the decision maker will gather more sample points about actions that are played more frequently.

As in Osborne and Rubinstein (1998), we assume that players draw naive frequentist inferences from their sample — that is, they treat the sample average as a predictor of the outcome they will get from each action, neglecting sampling error. This kind of over-inference from finite samples is related to the phenomenon that Tversky and Kahneman (1971) called "the law of small numbers".

A *representative sampling equilibrium* (RSE) in an extensive-form game is a profile of behavioral strategies that are consistent with this sampling procedure. That is, at every information set, the probability that the player takes the action a is the probability that it will have the best performance in the relevant representative sample.

The representative-sample assumption can be taken literally, modeling a form of experimentation in which players *deliberately* ensure that the composition of their sample matches the relevant population, in the manner of political pollsters. Our favored interpretation, however, regards the representative sample assumption as a *modeling approximation* of more passive, observational learning. In the smartphone story described above, agents base their decision on a *random* sample of their peers. Directly modeling a random-sample procedure would be more realistic. However, it is rather in-tractable; representative sampling is an approximation that makes the model tractable while keeping the idea that frequently played actions are sampled more often.

In line with this modeling strategy, we also assume that the signal the decision maker gets about an action from a single sample point is a *normally distributed* variable, whose mean and variance are those of the equilibrium conditional outcome distribution associated with this action. In some settings (e.g., when consumption outcomes in a single-agent choice model are normally distributed), this normally assumption is not an approximation but follows automatically from the model's primitives. In others (e.g., in a game with binary outcomes), normality is an approximation that addresses the problem that $n \cdot q(a)$ need not be an integer.

The modeling approximations of representative samples and normal variables constitute a methodological contribution of this paper: They enable tractable analysis of sampling-based equilibrium behavior in a variety of complex environments. Moreover, as we will see, non-trivial equilibrium effects persist even when n takes values for which these approximations are relatively accurate.

The basic insight of this paper is that since the sample size of each action depends on its popularity in the relevant population, it is endogenous and thus gives rise to a novel equilibrium effect. Even in a *single-agent* decision problem, the evaluation of a given action a will depend on its choice frequency $q(a)$, because the frequency affects the *variance* of the sample's outcome distribution; this variance in turn affects the probability with which the agent chooses a, which in equilibrium coincides with $q(a)$. This equilibrium effect is new to the literature on sampling-based solution concepts.

Revisiting our smartphone example, assume that smartphone A is objectively inferior to smartphone B. If a consumer were able to fully assess the brands' quality before making her purchase decision, she would choose B. However, when she bases her decision on a representative sample of her peers, she may decide to purchase A due to sampling error. Since A is objectively inferior, it is less likely to be chosen and therefore the sample will contain fewer observations of A users. As a result, the consumer's estimate of the quality of A will be noisier. Since a noisy assessment favors an objectively inferior alterative, it introduces an equilibrium effect that magnifies the choice frequencies of objectively inferior actions, compared with the choice frequencies under a uniform sample.

The observation that naive inference from representative samples introduces an equilibrium force that favors inferior alternatives is a key message of this paper. We explore its ramifications in various settings. In Sections 2 and 3 we present and analyze a simple model of *binary choice*, in which a consumer's underlying objective valuation of actions is a function of her private type. The same alternative is objectively superior for all consumer types. Thus, the only difference between types is in the intensity of this objective preference. We assume that the decision maker's representative sample is drawn from a population of consumers who are "similar" in the sense that they belong to the same *category* of types. We define RSE for this environment and obtain existence, uniqueness, and monotonicity results for RSE in this binary-choice model.

The basic insight described above implies that not only does the representative sample assumption increase the equilibrium frequency of the inferior alternative relative to the rational or uniform-sample cases, but the *rate* with which this frequency vanishes with n is extremely slow. We also carry out comparative statics with respect to the *coarseness* of the categorization of consumer types. We show that when the objective payoff difference between the two alternatives is not too large, a finer categorization leads to a *higher* overall equilibrium probability of choosing the objectively inferior alternative.

In Section 4, we present the more general formulation of RSE for games, and illustrate it with the one-shot Prisoner's Dilemma. Finally, in Section 5 we extend the representative-sample idea to encompass the *situation* in which the agent takes her action. While our main model assumes that players' total sample size is n for every information set, here we assume that it is proportional to the information set's equilibrium frequency: More frequent situations generate more sample points. This version is even closer to the idea of samples that are drawn from passive observation as opposed to active experimentation. It also introduces an additional layer of sample-size endogeneity, because the frequency of information sets is determined in equilibrium. We illustrate this difference with an infinite-horizon trust game and show how the extended version leads to endogenous patterns of reciprocity, which are impossible under the original concept.

Related literature

As mentioned above, this paper builds on a literature that incorporates learning from finite samples into the definition of equilibrium concepts in games. Osborne and Rubinstein (1998) introduced the concept of $S(K)$ equilibrium, in which each player samples each available strategy K (independent) times and chooses the best-performing strategy in her sample. Osborne and Rubinstein (2003) study a variant on this concept (in the context of a voting model), in which each player best-replies to a finite sample drawn from her opponents' strategies. Spiegler (2006a,b) studied price competition models in which consumers evaluate products using the $S(K)$ procedure.

Osborne and Rubinstein (1998,2003) assumed that players regard their sample as a noiseless estimate of the distribution from which it is drawn. This is what we referred to as "naive frequentist" inference, which this paper assumes as well. Salant and Cherry (2020) extended the sampling-based equilibrium approach to a more general class of statistical inference procedures, and proposed Bernstein polynomials as a tool for analyzing equilibria in certain classes of games. Unlike the present paper, Salant and Cherry (2020) maintained Osborne and Rubinstein's assumption that sample size is an exogenous parameter.

Goncalves (2020) formulated an equilibrium concept for games, based on a sequential sampling procedure. Each player has a prior distribution over the opponents' strategies, and she uses rational sequential sampling to gather more accurate information about them. The player stops sampling before she attains certainty, due to sampling costs; this is what generates random equilibrium behavior.

Our model is also related to the literatures on word-of-mouth learning (e.g., Ellison and Fudenberg (1995)) and the role of homophily in learning in social networks (e.g., Golub and Jackson (2012)). Unlike this paper, both literatures involve explicitly dynamic models.

2 A Single-Agent Binary Choice Model

An agent is facing a choice between two alternatives, denoted A and B. The agent's type is $t \in T$, where $T \subset \mathbb{R}$ is a finite set. Let $\mu \in \Delta(T)$ represent a distribution over types in a large population of agents facing the same choice problem. Denote the fraction of type t in the population by μ_t. The agent's objective expected payoff from choosing an alternative $z \in \{A, B\}$ given her type $t \in T$ is $u(z, t)$.

Let Π be a partition of T, where $\Pi(t)$ denotes the partition cell that includes t. For some of our results, we will assume that Π consists of "intervals" — i.e., if $\Pi(t) = \Pi(t')$ and $t < t'' < t'$, then $\Pi(t'') = \Pi(t)$. In this case, we refer to Π as an *interval partition.*

Let $q_t(z)$ be the probability that agents of type t choose z. The average frequency of choosing z among types in some set $S \subseteq T$ is

$$\bar{q}_S(z) = \frac{\sum_{t \in S} \mu_t q_t(z)}{\sum_{t \in S} \mu_t} \tag{1}$$

We will often use the abbreviated notation $q_t = q_t(B)$ and $\bar{q}_S = \bar{q}_S(B)$.

In our model, q_t is a consequence of agents' attempt to learn their payoffs from samples. An agent's total sample size is a positive integer n. The agent's estimate of $u(z,t)$ is independently and normally distributed as follows:

$$\hat{u}(z,t) \sim N\left(u(z,t), \frac{\sigma^2}{n\bar{q}_{\Pi(t)}(z)}\right) \tag{2}$$

where $\sigma^2 > 0$ is the payoff variance of a sample point from any alternative.

Definition 1 *A profile $(q_t)_{t\in T}$ is a **representative-sampling equilibrium (RSE)** if for every $t \in T$,*

$$q_t = \Pr(\hat{u}(B,t) > \hat{u}(A,t))$$

where this probability is calculated according to (2).

The idea behind this formulation is as follows. Before choosing an action, an agent of type t samples the payoff realizations of each alternative. The alternatives' representation in her sample matches their choice frequencies among the types in $\Pi(t)$.

By the assumption that $\hat{u}(A,t)$ and $\hat{u}(B,t)$ are independent normal variables,

$$\hat{u}(B,t) - \hat{u}(A,t) \sim N\left(u(B,t) - u(A,t), \frac{\sigma^2}{n\bar{q}_{\Pi(t)}(A)\bar{q}_{\Pi(t)}(B)}\right)$$

From now on, we identify t with the mean of this distribution — i.e.,

$$t = u(B,t) - u(A,t)$$

such that t measures the agent's underlying intrinsic preference for B over A.

Consequently, the equilibrium condition can be rewritten as

$$q_t = Pr\left[N\left(0, \frac{\sigma^2}{n(1-\bar{q}_{\Pi(t)})\bar{q}_{\Pi(t)}}\right) < t\right]$$

or, equivalently,

$$q_t = \Phi\left(\frac{t}{\sigma}\sqrt{n\bar{q}_{\Pi(t)}(1-\bar{q}_{\Pi(t)})}\right) \tag{3}$$

where Φ is the *cdf* of the standard normal distribution (we will use this notation consistently throughout the paper).

We will use (3) as our working definition of RSE in Section 3. This formulation highlights the novel equilibrium force introduced by representative sampling. The empirical choice frequencies given by $\bar{q}_{\Pi(t)}$ affect the variance of type-t agents' estimate of the payoff difference between the two alternatives, and therefore the agent's choice probabilities — which in turn affects the value of $\bar{q}_{\Pi(t)}$. More specifically, when the average choice distribution is more skewed (i.e., when $\bar{q}_{\Pi(t)}$ is closer to 1), the variance of $\hat{u}(B,t) - \hat{u}(A,t)$ rises, and this introduces an equilibrium counter-force toward a less skewed distribution.[1]

As can be seen from (3), the value of σ can be *normalized* to 1 without loss of generality (because we can rescale t). Therefore, from now on we set $\sigma = 1$.

We conclude this section with comments on the model's interpretation.

The interpretation of Π

One interpretation of this partition is that it captures coarse sample data. The agent would like to learn the outcome of choices by other agents who are *just like him*. However, she may lack sufficiently detailed information about the characteristics of other agents in the sample. Therefore, she settles for data about agents who share with him the characteristics that do appear in the dataset.

An alternative interpretation is that Π represents a particular word-of-mouth learning environment. Agents learn from the experiences of other,

[1]When $\bar{q}_{\Pi(t)}(z) = 0$, $\hat{u}(z,t)$ (as given by (2)) is ill-defined because it involves an infinite variance. We treat $N(0,\infty)$ as a well-defined distribution satisfying $\Pr(x \le c) = \frac{1}{2}$ for every c. Thus, the equivalent definition of RSE given by (3) is legitimate even when $\bar{q}_{\Pi(t)}(z) = 0$.

socially linked agents. The partition corresponds to a particular social network that consists of isolated cliques. When Π is an interval partition, a finer partition corresponds to a larger degree of homophily.

Naive extrapolation

The agent in our model takes sample outcomes at face value. That is, she regards the sample average $\hat{u}(z,t)$ as an accurate representation of her underlying average payoff from choosing z. Thus, she ignores both sampling error and the coarseness of her sample pool. In this sense, our agent naively extrapolates a belief from her sample.

The normality assumption

The assumption that $\hat{u}(z,t)$ is normally distributed can be interpreted literally — i.e., every sample point is objectively drawn from a normal distribution. Alternatively, the assumption can be viewed as a modeling approximation. That is, the objective distribution of each sample point is not necessarily normal, yet the central limit theorem allows us to approximate the distribution of the sample average by a suitable normal distribution. This alternative interpretation will be more pertinent in later sections.

3 Analysis

Our first result establishes equilibrium existence.

Proposition 1 *An RSE exists.*

Proof. Equation (3) defines a fixed point of a continuous mapping from $[0,1]^{|T|}$ to itself. It exists, by Brouwer's fixed-point theorem. ■

Comparison with a uniform sample

From the reduced definition of RSE given by (3), it is immediate that in any RSE, the agent chooses the intrinsically superior alternative with probability greater than $\frac{1}{2}$. It is not surprising that due to sampling errors, the inferior alternative is also chosen with positive probability. Yet, how big is this effect? A central theme of this paper is that the representative sample assumption gives rise to a force that favors the inferior alternative.

To see why, compare (3) with the case of a *uniform sample*, where each alternative is sampled $\frac{n}{2}$ times. In this case, the probability of choosing B is given by

$$r_t = \Phi\left(\frac{t\sqrt{n}}{2}\right) \tag{4}$$

This formula has two noteworthy features. First, it lacks the equilibrium effect that arises from the representative sample assumption. Second, since $\sqrt{q(1-q)} \leq \frac{1}{2}$ for any $q \in [0,1]$, r_t assigns higher probability to the favored alternative than any RSE value of q_t, for any type t. We will pursue this comparison further in Section 3.1.

The following collection of results restricts attention to the case in which B is the intrinsically superior alternative.

Proposition 2 *Assume $t > 0$ for every $t \in T$. Then, there is a unique RSE.*

Proof. Assume towards contradiction that $q = (q_t)_{t\in T}, q' = (q'_t)_{t\in T}$ are both RSE solutions and $q \neq q'$. Let t satisfy $q_t \neq q'_t$. Then, by (3), $\bar{q}'_{\Pi(t)} \neq \bar{q}_{\Pi(t)}$. Assume w.l.o.g that $\bar{q}_{\Pi(t)} > \bar{q}'_{\Pi(t)}$. Since $t > 0$ for every $t \in T$, we have $q_t, q'_t > \frac{1}{2}$ for every $t \in T$ and hence $\bar{q}_{\Pi(t)} > \bar{q}'_{\Pi(t)} > \frac{1}{2}$. This implies $\bar{q}_{\Pi(t)}(1-\bar{q}_{\Pi(t)}) < \bar{q}'_{\Pi(t)}(1-\bar{q}'_{\Pi(t)})$. Thus, for all $t \in \Pi(t)$,

$$q_t = \Phi\left(t\sqrt{n\bar{q}_{\Pi(t)}(1-\bar{q}_{\Pi(t)})}\right) < \Phi\left(t\sqrt{n\bar{q}'_{\Pi(t)}(1-\bar{q}'_{\Pi(t)})}\right) = q'_t$$

Hence,

$$\bar{q}_{\Pi(t)} = \frac{\sum_{t\in\Pi(t)} \mu_t q_t(z)}{\sum_{t\in\Pi(t)} \mu_t} < \frac{\sum_{t\in\Pi(t)} \mu_t q'_t(z)}{\sum_{t\in\Pi(t)} \mu_t} = \bar{q}'_{\Pi(t)}$$

a contradiction. ■

The next result establishes monotonicity of $\bar{q}_\pi$ when Π is an interval partition. Given any two cells $\pi, \pi' \in \Pi$, write $\pi' \succ \pi$ if and only if $t' > t$ for every $t \in \pi$, $t' \in \pi'$.

Proposition 3 *Suppose* Π *is an interval partition. Then, in equilibrium,* $\pi \succ \pi'$ *implies* $\bar{q}_\pi > \bar{q}_{\pi'}$.

Proof. Suppose that $\pi \succ \pi'$, and assume that $\bar{q}_{\pi'} > \bar{q}_\pi$. As we already saw, since $t > 0$ for every $t \in T$, $q_t > \frac{1}{2}$ for every t in RSE, and therefore $\bar{q}_\pi > \frac{1}{2}$. It follows that $\bar{q}_\pi(1-\bar{q}_\pi) > \bar{q}_{\pi'}(1-\bar{q}_{\pi'})$. Since $t > t'$ for every $t \in \pi$, $t' \in \pi'$, it follows from (3) that $q_t > q_{t'}$ in RSE for every $t \in \pi$, $t' \in \pi'$, hence $\bar{q}_\pi > \bar{q}_{\pi'}$, a contradiction. ■

Note that the monotonicity result applies to cells of the interval partition Π, but not necessarily to individual types. In particular, it is possible that $t' > t$ and yet $q_{t'} < q_t$ in the unique RSE. To see why, note that in RSE, two opposing forces shape choice probabilities. On one hand, a higher type (which represents a greater underlying taste for B) is a force that increases the probability of choosing this alternative. On the other hand, suppose that $\Pi(t') \succ \Pi(t)$ and t' is at the lower end of its cell while t is at the upper end of its cell. Then, t' shares its cell with higher types that imply a high $\bar{q}_{\Pi(t')}$, whereas t shares its cell with lower types that imply a low $\bar{q}_{\Pi(t)}$. As a result, the sample size for alternative A will be smaller for type t', which implies a noisy estimate of the payoff difference between the two alternatives. This force favors the inferior alternative A, and therefore lowers the probability of choosing B for t', relative to t. The net effect of these two forces is ambiguous.

However, *within* each cell, monotonicity holds.

Remark 1 *Let* q *be an RSE. If* $\Pi(t') = \Pi(t)$ *and* $t' > t$*, then,* $q_{t'} > q_t$.

This is an immediate consequence of (3). It holds for any partition (not only interval partitions).

3.1 Convergence Properties

Consider the case of a perfectly fine partition, $\Pi(t) \equiv \{t\}$. In this sub-section, we will use $q_t(n)$ to denote the RSE for type t and the sample size n, in order to highlight the role of n. It is uniquely given by

$$q_t(n) = \Phi\left(t\sqrt{nq_t(n)(1-q_t(n))}\right) \tag{5}$$

Our task is to analyze the asymptotic properties of $q_t(n)$, especially in comparison with the uniform sample case.

First, observe that $q_t(n)$ increases with n. The reason is that an increase in n is formally equivalent to an increase in t, which enables us to apply Proposition 3.

Proposition 4 $\lim_{n\to\infty} q_t(n) = 1$.

Proof. Assume the contrary — i.e., there exists $q^* < 1$ such that for every $n > 0$, there exists $n' > n$ such that $q_t(n') < q^*$. Recall that $q_t(n') > \frac{1}{2}$. Therefore, for all such n',

$$q_t(n')(1-q_t(n')) > q^*(1-q^*)$$

Consequently, $\sqrt{n'q_t(n')(1-q_t(n'))}$ diverges with n', which implies that, from some point onward,

$$\Phi\left(t\sqrt{n'q_t(n')(1-q_t(n'))}\right) > q^*$$

a contradiction. ■

Although the equilibrium effect of representative sampling does not stop choice behavior from converging to the rational-choice benchmark when n is large, it considerably slows down the convergence process, as the next result demonstrates. In particular, we compare the rate of convergence of $q_t(n)$

with the corresponding solution of the uniform-sample solution $r_t(n)$ given by (4) for $\sigma = 1$, which we can written as

$$r_t(n) = \Phi\left(\frac{t}{2}n^{\frac{1}{2}}\right)$$

We already showed that $r_t(n) > q_t(n)$ for every $t, n > 0$. We will now show $q_t(n)$ increases much more slowly than $r_t(n)$. For convenience, we fix $t = 1$; this is without loss of generality.

Proposition 5 *For every $k > 0$, there exists $n(k)$ such that for every integer $n \geq n(k)$,*

$$q_1(n) \leq \Phi(\frac{1}{2}n^k)$$

Proof. We will prove that for all $k > 0$,

$$q_1(n) \leq \Phi(n^k)$$

from some $n(k)$ onward. The general claim follows immediately with a suitable change of $n(k)$.

Let $n, k > 0$. Denote $x = q_1(n)$. That is, x is the unique solution to

$$x = \Phi\left(\sqrt{nx(1-x)}\right)$$

Assume $x > \Phi(n^k)$. Since Φ is monotonically increasing, $\sqrt{nx(1-x)} > n^k$ or, equivalently,

$$x(1-x) > n^{2k-1} \tag{6}$$

The contradiction is immediate for $k \geq \frac{1}{2}$. Henceforth, we assume $k < \frac{1}{2}$.

Let $f(x) = x(1-x)$. The function f is invertible for $x \in [\frac{1}{2}, 1]$ with $f^{-1} : [0, \frac{1}{4}] \rightarrow [\frac{1}{2}, 1]$ given by $f^{-1}(x) = \frac{1+\sqrt{1-4x}}{2}$. The inequality (6) implies

$0 < n^{2k-1} < \frac{1}{4}$ and, since f is strictly decreasing, also implies,

$$x < f^{-1}(n^{2k-1}) = \frac{1 + \sqrt{1 - 4n^{2k-1}}}{2}$$

Thus,

$$\Phi(n^k) < x < \frac{1 + \sqrt{1 - 4n^{2k-1}}}{2}$$

Hence, to reach a contradiction, it suffices to show that from some n onward,

$$\Phi(n^k) \geq \frac{1 + \sqrt{1 - 4n^{2k-1}}}{2}$$

By the Chernoff bound for the normal distribution (e.g., see Boucheron et al. (2013)),

$$1 - \Phi(x) \leq e^{-\frac{x^2}{2}} \tag{7}$$

for all $x > 0$. Thus,

$$\phi(n^k) \geq 1 - e^{-\frac{n^{2k}}{2}}$$

Hence, it suffices to prove

$$e^{-\frac{n^{2k}}{2}} \leq \frac{1 - \sqrt{1 - 4n^{2k-1}}}{2} \tag{8}$$

for sufficiently large n. See Claim 2 in the Appendix for this proof. ■

In the uniform-sample case, $r_t(n)$ increases with n like $\Phi(\sqrt{n})$. By comparison, in the representative sample case, $q_t(n)$ increases with n more slowly than $\Phi(n^k)$ for *any* k, however small (and in particular, smaller than $\frac{1}{2}$). Thus, the equilibrium forces introduced by representative sampling have a qualitative effect on the agent's choice behavior, even when n is large.

Figure 1 illustrates this comparison for $t = 1$. Figure 1(a) focuses on the range $n < 100$, while Figure 1(b) zooms out to $n < 500$ (and also describes $\Phi(\frac{1}{2}n^{1/4})$). As we can see, the uniform-case specification exhibits fast convergence — e.g., $r_1(30) \approx 0.997$. In contrast, the RSE prediction is

$q_1(30) \approx 0.925$, and convergence is very slow such that from around $n = 400$, $q_1(n) < \Phi(\frac{1}{2}n^{1/4})$.

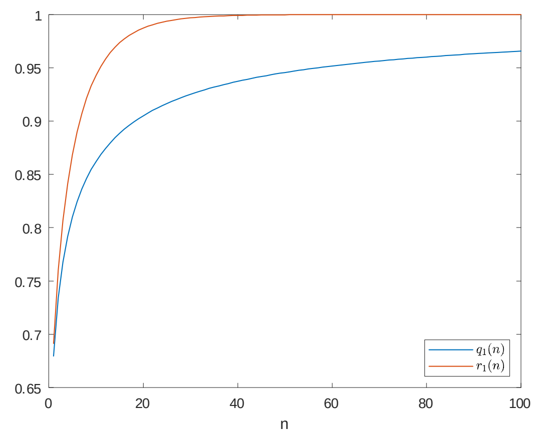

Figure 1(a)

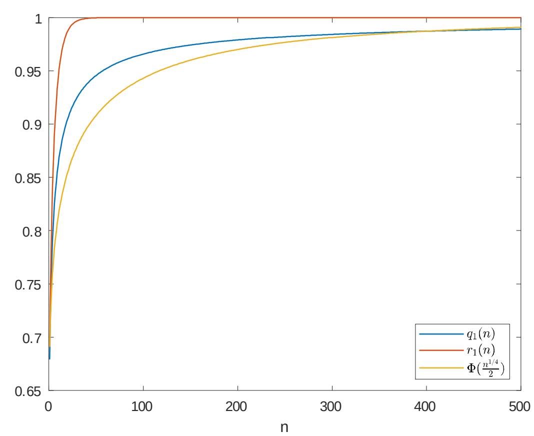

Figure 1(b)

3.2 The Effect of Partition Coarseness

We now turn to the question of how the coarseness of the partition Π affects the agent's behavior. First, we analyze the effect of splitting a partition cell into multiple sub-cells on the average behavior of types in the various sub-cells.

Proposition 6 *Consider two partitions Π and Π', such that Π' refines some cell T^* into a collection of sub-cells $\{T^1, ..., T^m\}$. Let q and q' be the RSE under Π and Π'. Then:*

(i) If $\bar{q}_{T^k} > \bar{q}_{T^*}$, then $\bar{q}'_{T^k} < \bar{q}_{T^k}$.

(ii) If $\bar{q}_{T^k} < \bar{q}_{T^*}$, then $\bar{q}'_{T^k} > \bar{q}_{T^k}$.

Proof. We prove part (i); the proof of part (ii) follows the same logic. Suppose $\bar{q}_{T^k} > \bar{q}_{T^*}$ for some $k = 1, ..., m$. Then, since both quantities are above $\frac{1}{2}$,

$$\bar{q}_{T^k}(1 - \bar{q}_{T^k}) < \bar{q}_{T^*}(1 - \bar{q}_{T^*})$$

By (3),

$$q_t = \Phi\left(t\sqrt{n\bar{q}_{T^*}(1 - \bar{q}_{T^*})}\right)$$

for every $t \in T^*$. Therefore, since Φ is an increasing function,

$$q_t > \Phi\left(t\sqrt{n\bar{q}_{T^k}(1 - \bar{q}_{T^k})}\right)$$

for every $t \in T^*$. Taking an average over $t \in T^k$ w.r.t the conditional type distribution given T^k, we obtain

$$\bar{q}_{T^k} - \sum_{t \in T^k} \frac{\mu_t}{\sum_{t \in T^k} \mu_t} \Phi\left(t\sqrt{n\bar{q}_{T^k}(1 - \bar{q}_{T^k})}\right) > 0 \quad (9)$$

By comparison, the definition of q' requires

$$\bar{q}'_{T^k} - \sum_{t \in T^k} \frac{\mu_t}{\sum_{t \in T^k} \mu_t} \Phi\left(t\sqrt{n\bar{q}'_{T^k}(1 - \bar{q}'_{T^k})}\right) = 0 \quad (10)$$

Since the L.H.S of (9)-(10) is an increasing function of a scalar variable ($\bar{q}_{T^k}$ in the inequality, $\bar{q}'_{T^k}$ in the equation), it follows that $\bar{q}'_{T^k} < \bar{q}_{T^k}$. ■

To understand this result, suppose that the original partition is fully coarse, and its refinement divides it into two groups. Suppose further that under the original coarse partition, the propensity to consume the superior alternative is above (below) the average in group 1 (2). What the result says is that after the refinement, the average probability of consuming the superior alternative in group 1 (2) decreases (increases). If we think of each cell in the refined partition as a "peer group", then the message of the result is that *increased homophily* (i.e., greater tendency to learn from similar types) leads to a *less skewed* choice distribution.

The intuition behind this result is that when members of group 1 stop learning from the choices of members of group 2, they have fewer sample points about the inferior product, which leads to a noisier assessment and therefore a lower probability of choosing the superior product.

While the above result holds for any partitional structure, in the remainder of the sub-section we restrict attention to interval partitions. Our result will make use of the following lemma (the proof is in the Appendix). Define the function

$$H(s,x) = \Phi\left(sx\right)$$

where $s, x > 0$.

Lemma 1 *If $s < 2$ and $x \in (0, \frac{1}{2})$, then H is supermodular.*

We now show that as long as the types in T are not too far away from zero, a finer partition leads to a higher overall probability of taking the inferior action A. Denote

$$\bar{q}(\Pi) = \sum_{t \in T} \mu_t q_t(\Pi)$$

where q_t^{Π} is the RSE probability that type t chooses B under the partition Π.

Proposition 7 *Suppose $t\sqrt{n} \in (0,2)$ for every $t \in T$. Consider two interval partitions Π and Π', such that Π' is a refinement of Π. Then, $\bar{q}(\Pi') < \bar{q}(\Pi)$.*

Proof. Take two interval partitions Π^c and Π^f, such that Π^f is a refinement of Π^c. For notational simplicity, let $q_t^f = q_t(\Pi^f)$ and $q_t^c = q_t(\Pi^c)$.

Consider some cell $T^* \in \Pi^c$. Denote

$$\alpha_t = \frac{\mu_t}{\sum_{s \in T^*} \mu_s}$$

Define

$$Q^c = \sum_{t \in T^*} \alpha_t q_t^c = \sum_{t \in T^*} \alpha_t \Phi\left(t\sqrt{Q^c\left(1 - Q^c\right)}\right)$$

This is the equilibrium probability of choosing B conditional on $t \in T^*$ under the partition Π^c.

Obviously, if T^* is also a cell in Π^f, then $q_t^c = q_t^f$ for every $t \in T^*$, hence $Q^C = Q^f$. We now turn to the non-degenerate case, in which Q^f strictly refines the cell T^* in Q^c.

Let β_π be the probability of $\pi \in \Pi^f$ conditional on $\pi \subset T^*$. Denote

$$\bar{q}_\pi = \sum_{s \in \pi} \frac{\alpha_s}{\beta_\pi} q_s^f$$

Define

$$Q^f = \sum_{t \in T^*} \alpha_t q_t^f = \sum_{t \in T^*} \alpha_t \Phi\left(t\sqrt{\bar{q}_{\Pi(t)}(1 - \bar{q}_{\Pi(t)})}\right)$$

This is the equilibrium probability of choosing B conditional on $t \in T^*$under Π^f.

Suppose that $Q^c \leq Q^f$. Then, since $\sqrt{q(1-q)}$ is strictly decreasing in $q > \frac{1}{2}$,

$$\sqrt{Q^c(1 - Q^c)} \geq \sqrt{Q^f(1 - Q^f)}$$

Since Φ is strictly increasing,

$$Q^c = \sum_{t \in T^*} \alpha_t \Phi\left(t\sqrt{Q^c\left(1 - Q^c\right)}\right) \geq \sum_{t \in T^*} \alpha_t \Phi\left(t\sqrt{Q^f\left(1 - Q^f\right)}\right)$$

Denote

$$x_\pi = \sqrt{\bar{q}_\pi(1 - \bar{q}_\pi)}$$

The expression $\sqrt{q(1-q)}$ is strictly concave in q. Therefore,

$$\sqrt{Q^f(1-Q^f)} = \sqrt{\left(\sum_{\pi \subset T^*} \beta_\pi \bar{q}_\pi\right)\left(1 - \sum_{\pi \subset T^*} \beta_\pi \bar{q}_\pi\right)} > \sum_{\pi \subset T^*} \beta_\pi \sqrt{\bar{q}_\pi(1-\bar{q}_\pi)} = \sum_{\pi \subset T^*} \beta_\pi x_\pi$$

Since Φ is strictly increasing,

$$\sum_{s \in T^*} \alpha_s \Phi(s\sqrt{Q^f(1-Q^f)}) > \sum_{s \in T^*} \alpha_s \Phi\left(s \sum_{\pi \subset T^*} \beta_\pi x_\pi\right) = \sum_{s \in T^*} \alpha_s H\left(s, \sum_{\pi \subset T^*} \beta_\pi x_\pi\right)$$

By concavity of H w.r.t its second argument,

$$H\left(s, \sum_{\pi \subset T^*} \beta_\pi x_\pi\right) > \sum_{\pi \subset T^*} \beta_\pi H(s, x_\pi)$$

for every s. Therefore,

$$\sum_{s \in T^*} \alpha_s H\left(s, \sum_{\pi \subset T^*} \beta_\pi x_\pi\right) > \sum_{s \in T^*} \sum_{\pi \subset T^*} \alpha_s \beta_\pi H(s, x_\pi)$$

Note that $x_\pi \in (0, \frac{1}{2})$ for every π, by the definition of x_π. Furthermore, by the monotonicity result, the cells in Π^f are ordered such that $\bar{q}_{\Pi^f(s)}$ is increasing in s, and hence $x_{\Pi^f(s)}$ is decreasing in s. By Lemma 1, H is supermodular when $s < 2$. Therefore,

$$\sum_{s\in T^*}\sum_{\pi\subset T^*}\alpha_s\beta_\pi H(s,x_\pi) > \sum_{s\in T^*}\alpha_s H(s,x_{\Pi^f(s)}) = \sum_{s\in T^*}\alpha_s\Phi\left(s\sqrt{\bar{q}_{\Pi(s)}(1-\bar{q}_{\Pi(s)})}\right) = Q^f$$

This inequality is a special case of a standard inequality from the literature on stochastic orderings — e.g., see Tchen (1980).[2] We have thus obtained $Q^c > Q^f$, a contradiction.

It follows that for every cell $T^* \in \Pi^c$, $Q^c \leq Q^f$, with a strict inequality for at least one cell. Therefore, $\bar{q}(\Pi^c) < \bar{q}(\Pi^f)$. ■

This result establishes that when the underlying payoff advantage of alternative B is sufficiently small, a finer partition leads to a higher probability of choice mistakes. Recall our two alternative interpretations of Π. Under the "coarse data" interpretation, the result means that finer data has an adverse effect on average choice quality. Under the "homophily" interpretation, the result means that increasing the homophily of the underlying social network that agents rely on for learning leads to poorer choice on average. The question of how the coarseness of Π affects average behavior for larger values of t remains open.

4 A General Formulation for Games

In this section, we extend the concept of RSE from single-agent decision problems to multi-agent games, and illustrate it with the one-shot Prisoner's Dilemma (in the next section, we apply RSE to an infinite-horizon game). For expositional purposes, we impose strong regularity conditions, avoid using the fully notated formalism of extensive-form games, and rely on verbal exposition whenever possible.

Consider a K-player extensive-form game. We use I_k to denote an information set at which player k moves. We use q_k to denote a behavioral

[2] We thank Meg Meyer for the reference.

strategy for player k, where $q_{k,I_k}(a)$ is the probability of playing action a that q_k induces at the information set I_k. We assume that any feasible profile $q = (q_k)_{k=1,\dots,K}$ of full-support strategies induces a well-defined probability distribution μ over all information sets in the game. Furthermore, for every information set I_k and for every action a that is feasible for player k at I_k, the distribution over player k's payoffs conditional on playing a at I_k is well-defined and has finite mean and variance. Finally, we assume that this payoff is additive and includes two components: one is known to player k at I_k, while the other is unknown and evaluated by sampling.

These assumptions allow a straightforward extension of RSE. Instead of a partition of the set of agent types, Π_k will represent a partition of player k's information sets. This is akin to the notion of analogy partitions in Jehiel (2005). At every such information set I_k, the player evaluates each feasible action a by sampling the payoff from playing a among the information sets in $\Pi_k(I_k)$. The composition of her sample is representative of the distribution induced by μ over $\Pi_k(I_k)$. Our assumption regarding the additive structure of payoffs allows us to accommodate games in which arriving at an information set reveals some payoff components (e.g., the price of a product) while the remaining components (e.g., product quality) remain unknown.

More formally, let

$$\bar{q}_{k,\pi} = \frac{\sum_{I_k \in \pi} \mu(I_k) q_{k,I_k}}{\sum_{I_k \in \pi} \mu(I_k)}$$

represent the average behavior of player k across all information sets in $\pi \in \Pi_k$. The player's estimated payoff from playing a at I_k is

$$\hat{u}_{k,I_k,q}(a) \sim N\left(m_{k,\Pi_k(I_k),q}(a) + t_{k,I_k}(a)\ ,\ \frac{\sigma^2_{k,\Pi_k(I_k),q}(a)}{n\bar{q}_{k,\Pi_k(I_k)}(a)} \right)$$

In this definition, $t_{k,I_k}(a)$ denotes the payoff component that player k knows at I_k (and therefore it does not depend on Π_k or q); and $m_{k,\Pi_k(I_k),q}(a)$ denotes the expectation of the player's payoff component that is unknown at I_k, conditional on being at any information set in $\Pi_k(I_k)$. As before, n denotes players' common sample size.

Definition 2 *A profile q of full-support strategies is an ε-RSE if for every player k, every information set I_k and every action a that is feasible for player k at I_k,*

$$q_{k,I_k}(a) = \varepsilon + (1-\varepsilon) \cdot \Pr\left[\hat{u}_{k,I_k,q}(a) > \hat{u}_{k,I_k,q}(a') \text{ for every other feasible } a'\right]$$

An RSE is a limit of ε-RSE for $\varepsilon \to 0$.

This extended definition of RSE assumes that what players evaluate by sampling is not their extensive-game strategies, but the actions that are feasible at any given information set. This is in the spirit of behavioral rather than mixed strategies in the classical theory of extensive-form games.

The extensive-form formalism introduces a new source of endogeneity that did not exist in the single-agent model. When the partition Π_k is not maximally fine, it consists of information sets whose probability according to μ is determined by players' strategies. In contrast, in the single-agent model μ was exogenous.

4.1 An Example: The Prisoner's Dilemma

Consider the following symmetric, simultaneous-move 2×2 game. There are two players, 1 and 2. The action set for each player is $\{0, 1\}$. Player i's payoff function is

$$u_i(a_1, a_2) = a_j - ca_i$$

where $j \neq i$ and $c < 1$. This is a standard specification of the Prisoner's Dilemma, where the strictly dominated action $a_i = 1$ corresponds to cooperation.

As in previous sections, our main interest in this sub-section is in the contrast between the predictions of RSE and the uniform-sample case.

Proposition 8 *The game has a unique symmetric RSE, where the probability of playing* $a = 0$ *is* $\Phi(c\sqrt{n})$.

Proof. Let q denote the RSE probability of $a = 0$. When a player draws a single sample point from an action a, she obtains the payoff $1 - ca$ with probability q and the payoff $-ca$ with probability $1 - q$. The normal distribution that shares the mean and variance with this random variable is

$$N\left(q - ca, q(1 - q)\right)$$

In RSE, the player samples $a = 0$ nq times and $a = 1$ $n(1 - q)$ times. Therefore, the player's estimated gain from playing $a = 0$ is

$$\hat{u}(0) - \hat{u}(1) \sim N\left(c, \frac{q(1-q)}{nq} + \frac{q(1-q)}{n(1-q)}\right) = N\left(c, \frac{1}{n}\right)$$

In RSE,

$$q = \Pr\left\{N\left(0, \frac{1}{n}\right) > -c\right\} = \Phi(c\sqrt{n})$$

This completes the proof. ■

Thus, RSE uniquely predicts a positive probability of cooperation, which is below $\frac{1}{2}$ and decreases with c and n. One might think that playing a strictly dominated action with positive probability is merely a consequence of sampling error. However, we will now see the crucial role that representative sampling plays in this result.

Specifically, compare our analysis with the uniform-sample case: a player's estimated gain from playing $a = 0$ is

$$\hat{u}(0) - \hat{u}(1) \sim N\left(c, \frac{2r(1-r)}{n} + \frac{2r(1-r)}{n}\right) = N\left(c, \frac{4r(1-r)}{n}\right)$$

where r is the probability that the player's opponent plays $a = 1$. The equilibrium condition for this uniform-sample variant is

$$r = \Pr\left\{N\left(0, \frac{4r(1-r)}{n}\right) > -c\right\} \tag{11}$$

Claim 1 *When $nc^2 > 8$, the unique solution of (11) is $r = 1$.*

Proof. The condition (11) can be rewritten as

$$r = \Phi\left(c\sqrt{\frac{n}{4r(1-r)}}\right)$$

Applying the Chernoff bound (7), we obtain

$$r = \Phi\left(c\sqrt{\frac{n}{4r(1-r)}}\right) \geq 1 - e^{-\frac{c^2 n}{8r(1-r)}}$$

Which is equivalent to

$$x \leq e^{-\frac{c^2 n}{8x(1-x)}}$$

Claim 3 in the Appendix establishes that when $nc^2 > 8$, this inequality fails for all $x \in (0, 1]$. ∎

This example demonstrates once again the key role of representative sampling in two-action decision problems — specifically, its enhancement of the perceived value of objectively inferior actions. In the Prisoner's Dilemma (as in any simultaneous-move game), the distribution of a single sample point for a player's action is given by the opponent's mixed strategy. As this strategy becomes more skewed in favor of the objectively superior action (defection), its variance vanishes and makes the player's assessment of the two actions more accurate. Under a uniform sample, this force eliminates the possibility of cooperative play when n is not too small. The representative-sample assumption introduces a counter-force that favors the objectively inferior action (cooperation) and therefore manages to sustain it with positive equilibrium probability for *any* value of n.

Comment. Arigapudi et al. (2021) study $S(K)$ equilibria in the Prisoner's Dilemma and their dynamic convergence properties. They show that for some range of values of K and the payoff parameters, cooperation can be part of a stable $S(K)$ equilibrium. However, if K is not small enough relative to the parameters that correspond to c in the present example, cooperation cannot be sustained in equilibrium. The uniform-sample version of the present model serves as a normal approximation of the analysis in Arigapudi et al. (2021), where $K = n/2$.

5 Situation-Dependent Sample Size (SDSS)

Our formulation of RSE for extensive-form games assumes that each player at any information set has a fixed "budget" of n sample points, which are allocated to the actions that are available at the information set according to their equilibrium frequencies in the relevant partition cell.

However, one could argue that the total number of sample points that a player has at some information set should reflect the frequency of the partition cell that contains it. If a class of information sets is rarely visited, then it is natural to assume that there will be few observations about it. In other words, the representative-sample idea may be extended to encompass not only actions but also the situations in which they are considered.

In this section, we explore the possible implications of this idea through a specific *infinite-horizon "trust" game* with an OLG flavor. We show that our previous definition of RSE implies a stationary cooperation pattern, whereas a variation that assumes situation-dependent sample sizes implies positive reciprocity in equilibrium.[3]

[3] It can be shown that a stationary equilibrium would exists under uniform sampling.

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

$$u(a_t, a_{t+1}) = a_{t+1} - ca_t$$

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

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. Thus, the set of relevant truncated histories is $H = \{0, 1\}^m$. For every truncated history $h = (a_{t-m}, ..., a_{t-1})$, (h, a) is a shorthand notation for the concatenated truncated history $(a_{t-m+1}, ..., a_{t-1}, a)$. A behavioral 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 truncated history h.

Benchmark: Nash equilibrium

As usual, this game has a Nash equilibrium in which every player chooses $a = 0$ after every history. This is the unique symmetric Nash equilibrium if we impose the following refinement: player t's equilibrium strategy conditions on an action in her truncated history only when she believes that this action affects the behavior of player $t+1$. 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 her 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 her behavior on a_{t-m^*}. By the refinement, she will not condition her own behavior on a_{t-m^*}, contradicting 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.

The game also has symmetric Nash equilibruia in which players cooperate. For instance, every f that satisfies $f(h,1) - f(h,0) = c$ is a symmetric Nash equilibrium, because players are always indifferent between the two actions. The function $f(h) = ca_{t-m} \cdots a_{t-1}$ is another symmetric Nash equilibrium that exhibits some cooperation. These equilibria violate the criterion that players condition on a past action only when they believe it is relevant for predicting future behavior.

For the remainder of the section, we assume $m = 2$ and present our results for this case only. Whether they can be extended to any $m > 2$ is an open problem.

5.1 RSE in the Trust Game

Let us see how RSE can be applied to this setting. A player's information set is the truncated history h. The partition is maximally fine: it coincides with the set of truncated histories H. Thus, when a player acts at the history h, she 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, she 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 player's assessment of the probability that a player's immediate successor cooperates after she herself plays $a = 1$ at h is

$$\hat{f}(h,1) \sim N\left(f(h,1), \frac{f(h,1)(1-f(h,1))}{nf(h)}\right) \tag{12}$$

Likewise, the player's assessment of the probability that a player's immediate successor cooperates after she herself plays $a = 0$ at h is

$$\hat{f}(h,0) \sim N\left(f(h,0), \frac{f(h,0)(1-f(h,0))}{n(1-f(h))}\right) \tag{13}$$

The player will weakly prefer to play $a = 1$ if and only if $\hat{f}(h,1) - \hat{f}(h,0) \geq c$. Therefore, in RSE, $f(h)$ is equal to

$$\Pr\left[N\left(f(h,1) - f(h,0), \frac{f(h,1)(1-f(h,1))}{nf(h)} + \frac{f(h,0)(1-f(h,0))}{n(1-f(h))}\right) > c\right]$$

Equivalently, this can be written as

$$f(h) = \Phi\left(\frac{\sqrt{n}(f(h,1) - f(h,0) - c)}{\sqrt{\frac{f(h,1)(1-f(h,1))}{f(h)} + \frac{f(h,0)(1-f(h,0))}{1-f(h)}}}\right) \tag{14}$$

Let us guess a stationary RSE, in which such $f(h) = b$ for every h. Then, equilibrium is unique and given by:

$$b = 1 - \Phi(c\sqrt{n})$$

This coincides with the RSE in the one-shot Prisoner's Dilemma that we studied in Section 4.1. We will proceed to show that it is the unique RSE in the present setting. The result makes use of the following lemma, which is proved in the Appendix.

Lemma 2 *Fix $f(h,1), f(h,0) \in (0,1)$. Then, there is a unique $f(h)$ that solves equation (14).*

Proposition 9 *Let $m = 2$. Then, the stationary equilibrium is the unique RSE.*

Proof. Lemma 2 establishes that there is a unique $f(h)$ solution to (14) for any given $f(h,1)$ and $f(h,0)$. By definition, these two objects do not depend on a_{t-m} (i.e., the earliest action in player t's truncated history). Then, this property necessarily extends to $f(h)$. When $m=2$, this means that in RSE, $f(h)$ is purely a function of the most recent action — i.e., $f(a_{t-2},a_{t-1})$ is constant in a_{t-2}.

Accordingly, let f_a denote the equilibrium probability that $a_t=1$ conditional on $a_{t-1}=a$. In addition, denote $x(a)=\hat{f}(a,1)-\hat{f}(a,0)$. Then,

$$x(1) \sim N\left(f_1-f_0, \frac{1}{n}\left(1-f_1+\frac{f_0(1-f_0)}{(1-f_1)}\right)\right)$$
$$x(0) \sim N\left(f_1-f_0, \frac{1}{n}\left(\frac{f_1(1-f_1)}{f_0}+f_0\right)\right)$$

Recall that

$$f_a = \Pr(\hat{f}(a,1)-\hat{f}(a,0) \geq c)$$

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,

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

Using the fact that $f_1>f_0>\frac{1}{2}$, we obtain

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

a contradiction.

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,

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

Since $f_1 < f_0 < \frac{1}{2}$, it follows that $f_1(1 - f_1) < f_0(1 - f_0)$, and we obtain a contradiction.

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 below c, $f_1 > f_0$ only if the variance of $x(1)$ is larger than the variance of $x(0)$. Therefore,

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

Since $f_0 < f_1 < \frac{1}{2}$, $f_0(1 - f_0) < f_1(1 - f_1)$, and we obtain a contradiction.

We have thus ruled out all possibilities of $f_1 \neq f_0$. ∎

Thus, RSE allows for cooperative behavior in the infinite-horizon trust game with $m = 2$, as a result of sampling errors — just as in the one-shot Prisoner's Dilemma of Section 4.1. However, it does not allow for any non-stationary patterns.

5.2 Emergent Reciprocity under SDSS

To introduce SDSS, note that a behavioral strategy f induces a discrete-time Markov process, in which the set of states is the set of truncated histories H. The probabilities of transition from $h \in H$ into the concatenated truncated histories $(h, 1)$ and $(h, 0)$ are $f(h)$ and $1 - f(h)$, respectively. If $f(h) \in (0, 1)$ for every h — i.e., f has full support — then the Markov process is irreducible and therefore has a unique invariant distribution over H, denoted α_f. Moreover, this distribution has full support.

For every $h \in H$ and $a \in \{0, 1\}$, define the following independently distributed, normal random variable:

$$\hat{f}(h,a) \sim N\left(f(h,a), \frac{f(h,a)(1-f(h,a))}{n\alpha_f(h,a)}\right) \tag{15}$$

This variable represents a player's estimate of the probability that the subsequent player will choose $a = 1$ following the truncated history (h, a). This is the same as (12)-(13), except that the number of observations about (h, a) is $n\alpha_f(h, a)$. This captures the idea that the representation of a situation in the sample is proportional to the frequency with which it is visited.

Definition 3 (Situation-dependent RSE) *A full-support strategy f is a situation-dependent RSE if, for every $h \in H$,*

$$f(h) = \Pr(\hat{f}(h,1) - \hat{f}(h,0) \geq c) \tag{16}$$

where $\hat{f}$ is defined by (15).

The following result shows that unlike the original definition of RSE, situation-dependent RSE involves non-stationary strategies. In particular, it implies positive reciprocity.

Proposition 10 *Let $m = 2$. In any situation-dependent RSE, $f(a_{t-2}, a_{t-1})$ is strictly increasing in a_{t-1}.*

Proof. Since f is effectively a function of the most recent action, we denote by f_a the probability that $a_{t+1} = 1$ conditional on $a_t = a$. In a similar vein, we use the notation α_h for $\alpha_f(h)$. Then, condition (16) can be written as

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

where

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

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

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

If $f_1 - f_0 \geq c$, then $f_1 > f_0$ and we are done. Now 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 representative samples, where representativeness extends to the truncated histories at which they evaluate actions. 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 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.

Unlike the reciprocity patterns admitted by Nash equilibrium, those that are implied by situation-dependent RSE are a lot more specific. We do not have an analytical proof that situation-dependent RSE is unique, but the numerical simulations we have conducted, presented in the following table, suggest that it is. They also give a sense of the magnitude of cooperation in the trust game for various values of c and n. Furthermore, RSE satisfies the criterion that players condition on a past action only when they believe it is relevant for predicting their opponent's behavior.

INSERT FIGURE

Situation-dependent RSE takes us further away from the "active experimentation" image behind $S(K)$ equilibrium and brings us closer to a sampling-based equilibrium concepts in which sample data is observational in nature.

6 Conclusion

This paper conveyed three basic ideas. First, it took the sampling-based equilibrium approach and modified its implicit "active experimentation" learning mode into a more passive format, which better fits situations in which players learn from observational data generated by their equilibrium behavior.

Second, the concept of RSE introduces two modeling approximations — representative samples and Gaussian approximations — which enhance the tractability of sampling-based equilibrium analysis and facilitate its extension to complex games.

Finally, the key equilibrium force that our paper highlighted is the effect of endogenous sample size on the variance of players' assessments of their actions. A skewed distribution over actions generates noisy estimates of their payoff differences, which favors the objectively inferior actions and thus moderates the distribution's skewness. This force drives new strategic effects, such as the slow convergence of equilibrium behavior toward the rational benchmark in binary choice models, or the play of dominated actions in 2×2 games.

Appendix: Missing Proofs

Claim 2 *For every sufficiently large n,*

$$e^{-\frac{n^{2k}}{2}} \leq \frac{1-\sqrt{1-4n^{2k-1}}}{2}$$

Proof. Define

$$h(n) = \frac{1-\sqrt{1-4n^{2k-1}}}{2} - e^{-\frac{n^{2k}}{2}}$$

Note that (since $k < \frac{1}{2}$) $lim_{n\to\infty} h(n) = 0$. Thus, it suffices to prove that there exists $n(k)$ such that for all $n \geq n(k)$, $h'(n) < 0$. This will imply $h(n) \geq 0$ for all $n \geq n(k)$ and thus that 8 holds for all such n. We have

$$h'(n) = \frac{(2k-1)n^{2k-2}}{\sqrt{1-4n^{2k-1}}} + kn^{2k-1}e^{-\frac{n^{2k}}{2}}$$

Therefore, $h'(n) < 0$ if and only if

$$\frac{e^{\frac{n^{2k}}{2}}}{n\sqrt{1-4n^{2k-1}}} > \frac{k}{1-2k}$$

Successive applications of L'Hôpital's rule imply

$$lim_{n\to\infty} \frac{e^{\frac{n^{2k}}{2}}}{n\sqrt{1-4n^{2k-1}}} = \infty$$

which completes the proof. ■

Proof of Lemma 1

Recall that

$$H(s,x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{sx} e^{-\frac{a^2}{2}} da$$

Let us calculate the cross derivative of H. First,

$$\frac{\partial H(s,x)}{\partial s} = \frac{1}{\sqrt{2\pi}} \cdot x \cdot e^{-\frac{1}{2}s^2x^2}$$

Now differentiate this expression w.r.t x:

$$\frac{1}{\sqrt{2\pi}} \left[e^{-\frac{1}{2}s^2x^2} - \frac{s^2}{2} \cdot 2x \cdot x \cdot e^{-\frac{1}{2}s^2x^2} \right]$$

When $x < \frac{1}{2}$, this expression is strictly positive whenever $s < 2$. ■

Claim 3 *If $x \in (0,1]$ and $nc^2 > 8$, then $x > e^{-\frac{c^2 n}{8x(1-x)}}$.*

Proof. Denote $t = c^2 n$ and define

$$f(x,t) = x - e^{\frac{-t}{8x(1-x)}}$$

Note that for all $x > 0$, $f(x,t)$ is increasing in t for $t > 0$. Thus, it suffices to prove that $f(x,8) > 0$ for all $x \in (0,1]$. For all such x we have $x > x(1-x) > 0$ and hence,

$$f(x,8) = x - e^{\frac{-1}{x(1-x)}} > x - e^{\frac{-1}{x}}$$

Denote $g(x) = x - e^{\frac{-1}{x}}$. We will prove $g(x) > 0$ for all $x > 0$. Note that $\lim_{x\to 0^+} g(x) = 0$. Thus, it suffices to prove $g'(x) > 0$ for all $x > 0$. We have,

$$g'(x) = 1 - \frac{1}{x^2} e^{\frac{-1}{x}}$$

Hence,

$$g'(x) > 0 \iff 2ln(x) > -\frac{1}{x}$$

We now show that $2ln(x) > -\frac{1}{x}$ for all $x > 0$. Consider $h(x) = 2ln(x) + \frac{1}{x}$. Note that $\lim_{x\to 0^+} h(x) = \lim_{x\to\infty} h(x) = \infty$. Thus, to prove $h(x) > 0$ for all $x > 0$, it remains to show that $h(x)$ has a unique interior extremum x^* (which must be a minimum) and that $h(x^*) > 0$. The first order condition is,

$$\frac{2}{x} - \frac{1}{x^2} = 0$$

The unique solution of this equation is $x^* = \frac{1}{2}$. Thus, for all $x > 0$, $h(x) > h(\frac{1}{2}) \approx 0.614 > 0$. ■

Proof of Lemma 2

Fix $f(h,1), f(h,0) \in (0,1)$. Denote

$$\begin{aligned} x &= f(h) \\ d &= \sqrt{n}(f(h,1) - f(h,0) - c) \\ a &= f(h,1)(1 - f(h,1)) \\ b &= f(h,0)(1 - f(h,0)) \end{aligned}$$

Then, equation (14) can be written as

$$x = \Phi\left(d\sqrt{\frac{x(1-x)}{a(1-x)+bx}}\right) \tag{17}$$

where $a, b \in (0, \frac{1}{4})$, d is any real number, and x is only allowed to get values in $[0,1]$.

When $d = 0$, $x = \frac{1}{2}$ trivially. Suppose $d > 0$ (the case of $d < 0$ is proved in the same manner). Then, the candidate solutions of (17) lie in $[\frac{1}{2}, 1]$. Moreover, the R.H.S is above $\frac{1}{2}$ (namely, above the L.H.S) at $x = \frac{1}{2}$ and takes the value 0 (namely, below the L.H.S) at $x = 1$. In addition, the function

$$h(x) = \frac{x(1-x)}{a(1-x)+bx}$$

is concave in x. Therefore, there is some $x^* \in [\frac{1}{2}, 1]$ such that h is increasing for $x < x^*$ and decreasing for $x > x^*$. Recall that the functions $\sqrt{z}$ and $\Phi(z)$ are strictly increasing and concave for $z > 0$. Therefore, the composite function

$$\Phi\left(d\sqrt{\frac{x(1-x)}{a(1-x)+bx}}\right)$$

is strictly increasing and concave for $x < x^*$, and decreasing (but not necessarily concave) for $x > x^*$. It follows that (17) has a unique solution in $(\frac{1}{2}, 1)$.

References

[1] Arigapudi, S., Y. Heller and I. Milchtaich (2021), Instability of defection in the prisoner's dilemma under best experienced payoff dynamic", Journal of Economic Theory 197, 105174.

[2] Boucheron, S., G. Lugosi and P. Massart (2013), Concentration inequalities: A nonasymptotic theory of independence, Oxford university press.

[3] Ellison, G. and D. Fudenberg (1995), Word-of-mouth communication and social learning, Quarterly Journal of Economics 110, 93-125.

[4] Fudenberg, D. and D. Levine (1998), The theory of learning in games, MIT press.

[5] Jehiel, P. (2005), Analogy-based expectation equilibrium, Journal of Economic theory 123, 81-104.

[6] Golub, B. and M. Jackson (2012), How homophily affects the speed of learning and best-response dynamics, Quarterly Journal of Economics 127, 1287-1338.

[7] Goncalves, D. (2020), Sequential sampling and equilibrium, mimeo.

[8] Osborne, M. and A. Rubinstein (1998), Games with procedurally rational players, American Economic Review, 834-847.

[9] Osborne, M and A. Rubinstein (2003), Sampling equilibrium with an application to strategic voting, Games and Economic Behavior 45, 434-441.

[10] Salant, Y. and J. Cherry (2020), Statistical inference in games, Econometrica 88, 1725-1752.

[11] Spiegler, R. (2006a), The market for quacks, Review of Economic Studies 73, 1113-1131.

[12] Spiegler, R. (2006b), Competition over agents with boundedly rational expectations, Theoretical Economics, 1, 207-231.

[13] Tchen, A. (1980), Inequalities for distributions with given marginals, Annals of Probability 8, 814-827.

[14] Tversky, A. and D. Kahneman (1971), Belief in the law of small numbers, Psychological bulletin 76, 105-110.