European Economic Review 189 (2026) 105462

Contents lists available at ScienceDirect

European Economic Review

journal homepage: www.elsevier.com/locate/eer

Procedural rationality, endogenous sampling, and the instability of defection ☆

Srinivas Arigapudi [a], Jonathan Newton [b],*

[a] Department of Economic Sciences, Indian Institute of Technology Kanpur, India
[b] Institute of Economic Research, Kyoto University, Japan

ARTICLE INFO

JEL classification:
C71
C73
D01

Keywords:
Procedural rationality
Imitation
Prisoner's dilemma
Sampling

ABSTRACT

We examine cooperation and defection in populations where individuals use both procedural rationality and imitation to choose their strategies. For a broad class of sampling-based choice procedures that encompasses both procedural rationality and imitation, we characterize the conditions under which universal defection is unstable. The results highlight that for cooperation to emerge and persist, the manner in which agents generate and share information can be as important as the underlying payoff incentives.

1. Introduction

It is standard in economics to assume that agents form beliefs about the uncertainties they face and choose strategies in response to these beliefs. However, in environments where players have limited information, alternative decision making procedures may be more appropriate. In this paper, we take two popular classes of decision making procedures and combine them. Under this new, broad class of rules we study the classic problem of the prisoner's dilemma and the evolution of cooperation. In particular, we study conditions under which defection, the strictly dominant strategy, is unstable in a population.

One belief-free decision making procedure is procedural rationality (Osborne and Rubinstein, 1998). To illustrate the choice procedure, we adapt an example from the cited paper. Alice has just moved to Bangalore and must choose a route to drive downtown to work. She has no information about which routes are typically congested. In such a scenario, it is plausible that she will experiment: testing different choices of route before settling on the one that performs best on average across her tests. Such procedurally rational decision making has been successfully applied to a variety of games (e.g. Spiegler, 2006; Sandholm et al., 2019; Sethi, 2021; Arigapudi et al., 2021).

Another belief-free decision making procedure is imitation. Alice asks neighbors about their commuting experiences, then adopts the commute that performs best on average across her neighbors' reports. There exists a large theoretical and experimental literature

☆ Arigapudi gratefully acknowledges financial support from the ANRF ARG MATRICS grant (ANRF/ARGM/2025/000809/QSS) and an Initiation grant at IIT Kanpur. Newton's work was supported by JSPS KAKENHI 22H00826. The program 'Decisions, Games, and Evolution' (ICTS/DGE2025/03), organized by the International Centre for Theoretical Sciences (ICTS) Bangalore played a pivotal role in the development of this project. We thank Ratul Lahkar and Minoru Osawa for helpful comments.

3 min video explainer: https://www.linkedin.com/posts/jonathan-newton-00867b84_another-new-working-paper-written-with-srinivas-activity-7396850004197318657-nZIR.

* Corresponding author.
E-mail addresses: arigapudi@iitk.ac.in (S. Arigapudi), newton@kier.kyoto-u.ac.jp (J. Newton).
URLs: https://sites.google.com/view/srinivasarigapudi/ (S. Arigapudi), https://jonathannewton.net (J. Newton).

https://doi.org/10.1016/j.euroecorev.2026.105462
Received 14 December 2025; Received in revised form 5 August 2026; Accepted 17 August 2026
Available online 24 August 2026

Table 1
Prisoner's dilemma. Two player game with strategies cooperate (C) and defect (D). Gain from unilateral defection is parameterized by $g > 0$, while loss from an opponent's defection is parameterized by $l > 0$.

	C	D
C	$1, 1$	$-l, 1+g$
D	$1+g, -l$	$0, 0$

on imitation (Alós-Ferrer and Schlag, 2009; Apesteguia et al., 2007), and it can be a relatively efficient way to make decisions (see, e.g. Schlag, 1998, 1999; Schipper, 2009; Duersch et al., 2012).

In this paper, we combine these decision making procedures into a single class. That is, we consider that Alice might combine direct information from experimentation (procedural rationality) with indirect information from neighbors (imitation). For example, she might follow a procedure by which she tries every route once and combines her information from this experience with information she learns from her neighbors. This unification of two classes of decision making procedure is the modeling contribution of the current paper.

Modeling Contributions. In the current paper, we combine procedural rationality and imitation in a population model. Agents in our model test each strategy m times to obtain direct information, then sample k strategies from the population to obtain indirect information, thus combining

(a) Dynamics of procedural rationality, known as sampling dynamics (Sethi, 2000) or best experienced payoff dynamics (Sandholm et al., 2020).
(b) Imitative dynamics in which the amount of information obtained about a strategy depends on its frequency in the population (W.H. Sandholm, 2010).

This hybridization is natural because both procedural rationality and imitation are belief-free procedures. Agents need not know the payoff structure of the game, but instead rely on realized payoffs from tested or observed strategies. In environments with limited information, it is therefore plausible that agents combine direct experimentation with indirect information from the experiences of others, rather than relying exclusively on one source of payoff information.

The prisoner's dilemma (Table 1) is famous in game theory, economics and biology. It is the workhorse that underpins models of public good provision (Hardin, 1968; Fehr and Gächter, 2000) and the broader evolution of cooperation (Bowles and Gintis, 2011; Hilbe et al., 2013; Glynatsi et al., 2024). An important topic in this literature is the study of conditions under which cooperation proliferates when rare. That is, starting from universal defection — the population state at which every player defects, if a small amount of cooperation enters the population, does it then increase its share, or does it shrink back to nothing? In other words, is universal defection unstable or stable?

We study conditions under which cooperation proliferates when rare under our large class of hybrid dynamics that nest both procedural rationality and imitation. These dynamics are summarized in Fig. 1. We find that results can qualitatively differ to those obtained under either procedural rationality alone (Arigapudi et al., 2021) or imitation alone.[1]

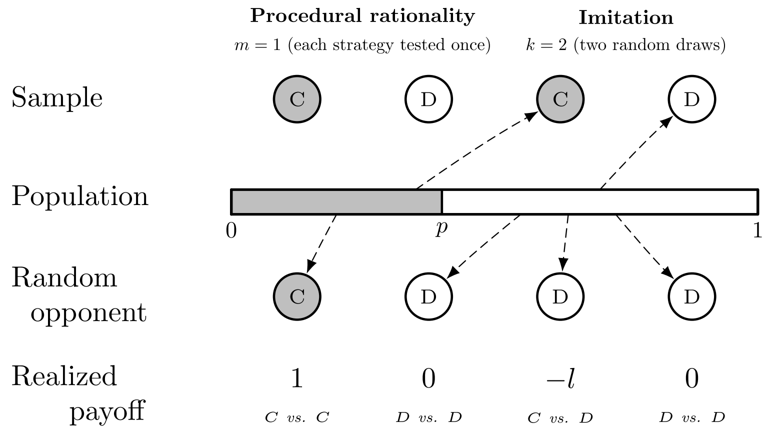
Finding 1. *Universal defection is stable under procedural rationality when each strategy is tested once ($m = 1$).*[2] *Universal defection is also stable under imitation when $k \geq 2$ strategies are sampled from the population.*[3] *However, under the hybrid procedure that combines these two sources of information ($m = 1,\ k \geq 1$), universal defection is unstable when gains from defection and losses from being defected against (g and l in Table 1) are sufficiently small.*

Before we describe our general characterization, it is worth considering why Finding 1 is surprising and subtle. Consider $m = k = 1$ and states close to universal defection where there is very little cooperation in the population.

(i) The procedural rationality component tests each of C and D once ($m = 1$).
(ii) The imitative component tests one strategy ($k = 1$) drawn from the population. This will almost always be D, as there is little C in the population.

Therefore, the hybrid procedure will, with high probability, test C once (directly) and D twice (once directly and once indirectly). It can be shown that if this were *always* the case, then universal defection would be stable. However, the small probability of C being the indirectly sampled strategy in step (ii) suffices to overturn the advantage of defection. Universal defection becomes unstable and cooperation proliferates when rare. Moreover, the resulting amount of cooperation is qualitatively significant: we obtain an almost globally stable state at which a considerable share (29.3%) of agents cooperate (Fig. 3).

[1] The dynamic for imitation on its own is monotone in the sense of Nachbar (1990), so the strictly dominated strategy C (cooperate) is eliminated over time.
[2] Osborne and Rubinstein (1998) call the procedure when each strategy is tested exactly once the $S(1)$ procedure. It is the primary focus of their exposition and analysis.
[3] When $m = 0$ and $k = 1$, every population state is a stationary state, so nothing happens.



- Average realized payoff from C is $\frac{1-l}{2}$. Average realized payoff from D is 0.
- Therefore C is chosen if and only if $l < 1$.

Fig. 1. Summary of choice procedure. The share of agents in the population currently playing C is given by p. The choice procedure combines two components. Procedural rationality directly tests strategies a given number of times against random opponents. Imitation indirectly tests strategies by randomly observing instances of them being played against random opponents. The tested strategy that gives the highest average realized payoff is then chosen, with tie breaking in favor of D.

In general, we derive necessary and sufficient conditions for universal defection to be unstable. These conditions are a combination of the amount of direct and indirect information used, as well as the payoff parameters of the game.

Finding 2. *Universal defection is unstable if and only if the loss from being defected against (l) is low and either*

(I) There is enough procedural rationality ($m \geq 2$); or
(II) There is a minimal amount of procedural rationality ($m = 1$), some amount of imitation ($k \geq 1$) and the gain from defection (g) is low.

In addition to the subtleties of the $m = 1$ case described above, Finding 2 describes instability arising from testing each strategy at least twice ($m \geq 2$). When cooperation is rare and is tested at least twice, the probability that at least one test of C is against an agent playing C exceeds the share of agents playing C in the population. When the payoff from a single such test outweighs the losses from the remaining tests of C against agents playing D, cooperation is chosen sufficiently often for universal defection to be unstable.

Finding 2 summarizes Theorems 1, 2, and 3 of the paper. It answers questions about competing effects. For example, it tells us that

$$\underset{\text{defection unstable}}{(m = 2,\ k = 0)} \quad , \quad \underset{\text{defection stable}}{(m = 0,\ k \geq 2)} \quad \text{and} \quad \underset{\text{defection unstable}}{(m = 2,\ k \geq 2)},$$

as well as implying Finding 1,

$$\underset{\text{defection stable}}{(m = 1,\ k = 0)} \quad \text{and} \quad \underset{\text{defection stable}}{(m = 0,\ k \geq 2)} \quad \text{yet} \quad \underset{\text{defection unstable}}{(m = 1,\ k \geq 1)}.$$

The above results are for a fixed amount of information. That is, m and k are fixed. A natural extension and robustness check is to consider situations in which the overall amount of information collected is not fixed. Specifically, we consider situations in which an agent wishes to test each strategy a minimum of m times and keeps sampling until this criterion is satisfied.

Finding 3. *Under sequential sampling until each strategy is tested a minimum of m times, universal defection is unstable if and only if the loss from being defected against (l) is low and $m \geq 2$.*

This finding summarizes Theorems 4 and 5 of the paper. The result is cleaner than that for fixed sample sizes. The reason for this is that, for $m = 1$ and population states close to universal defection, the probability of testing strategy C more than once is even lower than in the fixed sample size model. Indeed, it is low enough to remove the possibility of instability.[4]

[4] Writing p as the share of C in the population, the probability of testing C more than once is of order $\Theta(p)$ for the fixed sample size model, but of order $\Theta(p^2)$ under the minimum threshold model.

Overall, we provide a complete characterization of when cooperation proliferates when rare under our class of hybrid dynamics that spans procedural rationality, imitation, and the space in between. In doing so, we not only derive new results, but also obtain an understanding of the robustness of existing results. This will be discussed further in the Related Literature section.

The overall message of our results is the importance of testing strategies twice or more. Finding 2-I and Finding 3 show that as long as this is the case, and as long as the risk from cooperating (l) is not too high, cooperation will proliferate when rare. This occurs without any explicit altruism, via the invisible hand of sampling and testing possible strategies.[5]

Outline of the paper: Following a brief discussion of related literature and positioning of our results, Section 2 presents the model. Section 3 presents the results and Section 4 concludes. Proofs are in Appendix A.

Related literature & contribution

The version of procedural rationality studied here originates with Osborne and Rubinstein (1998) and Sethi (2000), and was later generalized by Sandholm et al. (2020). Procedurally rational learning dynamics have been applied to a wide range of strategic environments, including price competition with boundedly rational consumers (Spiegler, 2006), common-pool resources (Cárdenas et al., 2015), public good provision (Mantilla et al., 2018), centipede games (Sandholm et al., 2019), finitely repeated games (Sethi, 2021), the prisoner's dilemma (Arigapudi et al., 2021), the traveler's dilemma (Berkemer et al., 2023), the trust game (Arigapudi and Lahkar, 2024), the bilingual game (Arigapudi, 2024), the ultimatum game (Lahkar and Mahanta, 2025), and the hawk–dove game (Arigapudi and Heller, 2025). A central theme in this literature is that under procedurally rational learning dynamics, strict Nash equilibria can be unstable and non-Nash equilibria can be stable.[6]

This paper makes several contributions to this literature. We develop a unified framework that combines procedural rationality with imitative dynamics. Within this framework, we provide conditions under which cooperation can proliferate when rare, in the sense that the share of cooperators initially increases from population states close to universal defection, a central concern in the literature on the evolution of cooperation (see,e.g. Boyd et al., 2010). Our analysis demonstrates that results on stability under procedural rationality for $m = 2$ are robust to both the inclusion of additional information ($k \geq 1$, Finding 2) and to changes in the sampling procedure (Finding 3).[7] In contrast, the predictions of the procedural rationality literature for $m = 1$ are less robust (Finding 1), but remain valid in most circumstances (Finding 2-II).

A final methodological contribution of the current paper is to extend the framework to sequential sampling, where the overall quantity of information, and not only the strategies to which the information pertains, is endogenously determined. We prove results for this setting (Finding 3) that are similar, but not identical to, those for fixed m and k. Overall, the paper's perspective on combining sample selection, procedural rationality, and imitation should be applicable to other environments beyond the prisoner's dilemmas studied here.

From a broader perspective, our analysis adds to the literature on hybrid evolutionary dynamics. Notably, Sandholm (2005), W. Sandholm (2010) show that adding even a small amount of direct strategy evaluation (excess payoff dynamics and pairwise comparison dynamics respectively) to imitative dynamics can restore Nash stationarity. The present paper is similar in that hybridization can create or preserve important stability and instability properties.

2. Model

2.1. Population dynamics

There is a population of unit mass. Agents from this population are randomly matched to play the prisoner's dilemma game with payoffs shown in Table 1. Let $p(t) \in [0, 1]$ be the share of agents playing strategy C, and $1 - p(t)$ the share of agents playing strategy D at time t. These shares evolve over time $t \in \mathbb{R}_+$ as agents exit and are replaced by new agents entering the population. Agents exit the population at rate one per unit time, independently of their strategies. New agents also enter at rate one per unit time.

Each entering agent observes a nonempty sample of strategies, possibly with repetitions. For each sampled strategy, the agent also observes a realized payoff from playing that strategy against an independently drawn opponent from the current population

[5] To borrow from the lyrics of the song "Santa Claus Is Comin' to Town" by J. Fred Coots and Haven Gillespie (1934), we might say that making a list [of strategies] and checking it twice...actually makes agents nice [as in cooperative].

[6] A related literature considers *sampling best response dynamics*, where an agent best responds to a distribution over strategies inferred from a sample of k other agents. These dynamics were first studied by Sandholm (2001) and Osborne and Rubinstein (2003). Recent contributions include Oyama et al. (2015), Heller and Mohlin (2018), Salant and Cherry (2020), Sawa and Wu (2023), Arieli and Arigapudi (2024), Danenberg and Spiegler (2025), and Arigapudi et al. (2025). The critical distinction between procedurally rational learning and sampling best response dynamics lies in the treatment of samples: under procedural rationality, each strategy is tested against the strategies of independently drawn opponents, whereas under sampling best response dynamics, all strategies are evaluated against the same sample. Procedurally rational learning thus captures scenarios where agents lack knowledge of the payoff structure and rely only on experienced payoffs, while sampling best response captures scenarios where agents know the payoff structure but sample to learn about the play of other agents in the population.

[7] Arigapudi et al. (2021) show that under pure procedural rationality in a prisoner's dilemma game with $g, l < 1$ and $m = 2$, there is an almost globally stable state at which approximately 28% of agents cooperate.

in the game of Table 1. Precise mechanisms for generating these samples and payoffs will be described later: they may arise from agents testing strategies themselves, from observing the play and payoffs of others, or from a combination of both.[8]

For a given agent, we refer to the occurrences of strategy C in her sample as the **C**-sample and the occurrences of strategy D as the **D**-sample. Denote the size of her C-sample by N_c and the size of her D-sample by N_d. Given a sample and payoffs, the entering agent adopts a strategy as follows:

(I) If her sample is homogeneous in that it only contains C (respectively D), then the entering agent adopts strategy C (respectively D).
(II) If her sample is heterogeneous, then she adopts C if and only if the mean realized payoff for the C-sample is strictly greater than the mean realized payoff for the D-sample. Otherwise, she adopts strategy D.[9]

Example 1. Consider an entering agent who obtains a sample of size 3, with two occurrences of C and one occurrence of D. For each element of her sample, she observes a realized payoff obtained when the strategy is tested against an opposing strategy independently drawn according to $p(t)$. From the payoffs in Table 1, each C in her sample gives a payoff of 1 with probability $p(t)$ and a payoff of $-l$ with probability $1-p(t)$, and each D in her sample gives a payoff of $1+g$ with probability $p(t)$ and a payoff of 0 with probability $1-p(t)$. Consider the case in which her C-sample gives realized payoffs of 1 and $-l$ and her D-sample gives a realized payoff of 0. Then she adopts strategy C if and only if the mean payoff $(1-l)/2$ from her C-sample is strictly greater than the mean payoff of 0 from her D-sample. Otherwise, she adopts strategy D.

Given the decision making procedure described above, let $\Pr[C \mid p]$ denote the probability that an entering agent chooses C conditional on the population share of C being p. Writing $\dot{p}$ as the time derivative of p,

$$\dot{p}(t) = \Pr[C \mid p(t)] - p(t).$$

The time argument t will be frequently omitted and left implicit.

2.2. Sample selection

This paper considers different ways in which an agent's sample sizes N_c and N_d are determined. We first consider a fixed total sample size $N_c + N_d$.

Definition 1 (*Fixed Sample Size*).
An entering agent samples a baseline of $m \in \mathbb{Z}_{\geq 0}$ plays of each strategy, then augments this sample by observing k agents drawn uniformly from the population. In this case, we have

$$(N_c, N_d) = \big(m + K_c,\ m + k - K_c\big), \quad \Pr[K_c = k_c] = \binom{k}{k_c} p^{k_c}(1-p)^{k-k_c}.$$

We assume that $m + k \geq 1$, so that the entering agent has a nonempty sample.

It is possible that one of m or k is equal to zero. Indeed, these cases correspond to well known rules from the literature.

Example 2 (*Fixed Sample Size: Special Cases*).

- $m = 0,\ k \geq 1$. This is the classic case of simple random sampling.
- $m = 1,\ k = 0$. This is procedural rationality as described by Osborne and Rubinstein (1998). Each strategy is tested once, with the most successful being adopted.
- $m \geq 1,\ k = 0$. A generalization of procedural rationality is adoption according to best experienced payoffs, as in Sandholm et al. (2020). Each strategy is tested m times and the strategy giving the highest mean realized payoff is adopted.

In general, we can have both $m, k \geq 1$. For $k \geq 1$, the distribution over sample sizes (N_c, N_d) depends on the share p of cooperators in the population.

Another natural class of rules is threshold rules, under which the sample size is not fixed, but instead the agent keeps sampling until both C and D have been sampled a satisfactory number of times.

[8] Note that, unlike imitative rules that evaluate sampled strategies according to current *expected* payoffs, our choice procedures evaluate sampled strategies according to *realized* payoffs against independently drawn opponents. That is, there can be randomness in both sample selection and in the realized payoffs associated with the sample.

[9] Note that this implicitly assumes tie breaking in favor of D. This is not of qualitative importance to our results, but making some assumption is beneficial for clarity of exposition. Naturally, tie breaking towards D makes results on the instability of defection even more noteworthy.

Definition 2 (*Threshold*).

An entering agent draws agents uniformly from the population, stopping when there are at least $m \in \mathbb{N}$ strategies of each type in his sample. Thus,

$$\Pr\left[(N_c, N_d) = (m, m)\right] = \binom{2m}{m} p^m (1-p)^m,$$
$$\Pr\left[(N_c, N_d) = (m, k), k > m\right] = \binom{k+m-1}{m-1} p^m (1-p)^k,$$
$$\Pr\left[(N_c, N_d) = (k, m), k > m\right] = \binom{k+m-1}{m-1} p^k (1-p)^m.$$

Under threshold rules, not only does the distribution over sample sizes (N_c, N_d) depend on the share p of cooperators in the population, but the distribution over total sample sizes $N_c + N_d$ also depends on p.

Note that, at the boundary states $p = 0$ and $p = 1$, this stopping rule does not terminate. Therefore, we define $\Pr[C \mid p]$ for $p = 0, 1$ by the corresponding limit from the interior, for example $\Pr[C \mid 0] = \lim_{p \downarrow 0} \Pr[C \mid p]$.

2.3. Stationarity and stability

Before stating our results formally in the next section, we review some standard definitions. These provide equilibrium and stability concepts with which we can analyze our model. Consider the dynamical system from Section 2.1,

$$\dot{p} = \Pr[C \mid p] - p =: F(p). \tag{1}$$

A share of cooperators p^* in the population is a stationary state if it is a fixed point of (1). That is, if the share of cooperators among those exiting the population exactly equals the share of entering agents who choose to cooperate.

Definition 3. $p^* \in [0, 1]$ is a stationary state if $F(p^*) = 0$.

We are also interested in stability of stationary states under our dynamics. If we move the population slightly away from a stationary state p^*, does the process move away from p^*, stay close to p^*, or converge back to p^*? To consider such questions, we use standard concepts.

Definition 4. A stationary state p^* is

– Lyapunov stable if, for every neighborhood U of p^*, there is a neighborhood $V \subset U$ of p^* such that if $p(0) \in V$, then $p(t) \in U$ for all $t > 0$.
1. asymptotically stable if it is Lyapunov stable and there is some neighborhood U of p^* such that $p(0) \in U$ implies $\lim_{t\to\infty} p(t) = p^*$.
– unstable if it is not Lyapunov stable.
– almost globally stable if, for all $p(0) \in (0, 1)$, $\lim_{t\to\infty} p(t) = p^*$.

In words, Lyapunov stability of p^* means if the state is close enough to p^*, then it will never move far from p^*. Asymptotic stability means if the state is close enough to p^*, then it will converge to p^*. Almost global stability means that, starting from any interior state, the process will converge to p^*. These stability concepts are nested in terms of strength, with almost global stability implying asymptotic stability which implies Lyapunov stability.

3. Analysis

3.1. Fixed sample size

This section gives results for the fixed sample size selection procedures described in Definition 1. Results are arranged according to m, the baseline number of times each strategy, C and D, appears in the tested sample. We begin with $m = 0$, where there is no baseline, so the only data used by an agent when adopting her strategy comes from a simple random sample of size k. The case of $k = 1$ is trivial ($\dot{p} \equiv 0$ so the population state never changes), but included for completeness. For $k \geq 2$, universal defection ($p = 0$) is asymptotically stable [Fig. 2].

Theorem 1. *Consider simple random sampling,* $m = 0$, $k \geq 1$.

- *If* $k = 1$, *then every* $p \in [0, 1]$ *is a stationary state.*
- *If* $k \geq 2$, *then* $p = 0$ *is asymptotically stable.*

Naturally, it is possible to go beyond Theorem 1 and derive more granular results for specific values or ranges of l, g or k. Notably, for $k = 2$ and $k = 3$, it can be shown that $p = 0$ is almost globally stable for all l and g.

Next we consider the case of $m = 1$. When $k = 0$, this is the $S(1)$-equilibrium of Osborne and Rubinstein (1998) as studied by Sethi (2000). In this case, we know from Example 5 of the latter paper, or more generally from Arigapudi et al. (2021), that universal defection is globally stable. However, if $k \geq 1$, it turns out that universal defection is unstable if both the gains and losses from unilateral defection are low.

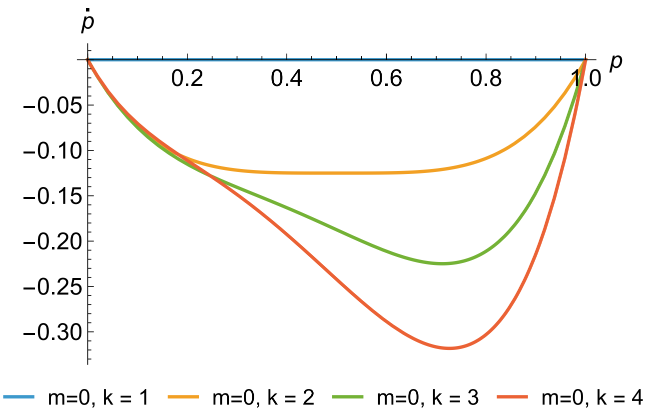


Fig. 2. Time derivative of p as a function of p. For $m = 0$, Theorem 1 shows that universal defection ($p = 0$) is asymptotically stable for $k \geq 2$, while for $k = 1$ every state is stationary. Here we illustrate $k = 1, 2, 3, 4$. For $k = 2, 3, 4$, take $l < \frac{1}{k-1}$ and $g > k - 1$. Under these parameters, for C to be chosen by an entrant with a mixed sample, at least one element of her C-sample must obtain payoff 1 and every element of her D-sample must obtain payoff 0.

Theorem 2. *Consider* $m = 1$.

- *If* $k \geq 1$, $g < k$, $l < 1$, *then* $p = 0$ *is unstable.*
- *Otherwise,* $p = 0$ *is asymptotically stable.*

Taking $k = 0$ as a comparison, let us consider how instability arises for $k = 1$. In the $k = 0$ case, we have $N_c = N_d = 1$. Given a share p of cooperators in the population, the probability of the C-sample having a higher average payoff than the D-sample is exactly $p(1-p)$, which is clearly less than p, therefore the process converges to $p = 0$. Letting $k = 1$, we have either $N_c = 2$ (w.p. p) or $N_d = 2$ (w.p. $1 - p$). There are two effects to consider. Firstly, when $N_c = 2$ and l is low, the probability of the C-sample having the higher average payoff becomes greater than p. This on its own is not enough to destabilize universal defection, as the event $N_c = 2$ is itself a rare event. In addition, a second effect is required: when $N_d = 2$ and g is low, the probability of the C-sample having the higher average payoff becomes $p(1 - p^2)$. This is greater than the $p(1 - p)$ we obtained under the $N_c = N_d = 1$ case, but still less than p. Thus, this second effect on its own does not induce instability. However, the combination of these two effects is enough to induce instability, leading to slow growth of p, with $\dot{p}$ of the order of p^2. See Fig. 3 for a graphical illustration of $\dot{p}$ under these effects.

Next we consider the case of $m \geq 2$. With this baseline, N_c is always at least two, and when l is low, for small values of p, the probability of the C-sample having higher average payoff than the D-sample is greater than p. The instability of universal defection under best experienced payoff dynamics (Arigapudi et al., 2021) with $m \geq 2$, $k = 0$ is preserved under fixed sample size procedures with additional random sampling ($k \geq 1$) [Figs. 4, 5].

Theorem 3. *Consider* $m \geq 2$.

- *If* $l < \frac{1}{m-1}$, *then* $p = 0$ *is unstable.*
- *If* $l \geq \frac{1}{m-1}$, *then* $p = 0$ *is asymptotically stable.*

Table 2 reports stable stationary states for $g = l = 0.2$ across selected values of m and k, computed numerically using Wolfram Research, Inc. (2026). The selected parameter values illustrate the different cases covered by Theorems 1, 2, and 3. The table also reveals several additional patterns. Large m eventually eliminates interior stationary states. When $l < 1/(m-1)$, increasing k causes the interior stable state to converge slowly towards zero. When $l \geq 1/(m-1)$, sufficiently large k can instead eliminate the interior stationary state altogether.

3.2. Sequential sampling to a threshold

This section gives results for the class of threshold sample selection procedures described in Definition 2. This is another plausible way to select a sample and thus of independent interest as well as acting as a robustness check for our fixed sample size results. In particular, proofs are more involved and due to unbounded sample sizes, we have to bound tail probabilities (Arratia and Gordon, 1989).

First we deal with a threshold of $m = 1$. That is, an entering agent will continue to sample until she has at least one C and at least one D in her tested sample. For low values of p, there will typically be several D's before the first C is encountered and the sample selection terminates. Thus, a typical sample will have $N_c = 1$, $N_d = k$. This is similar to the fixed sample size $m = 1$ case. However, results turn out to be cleaner, with universal defection always asymptotically stable.

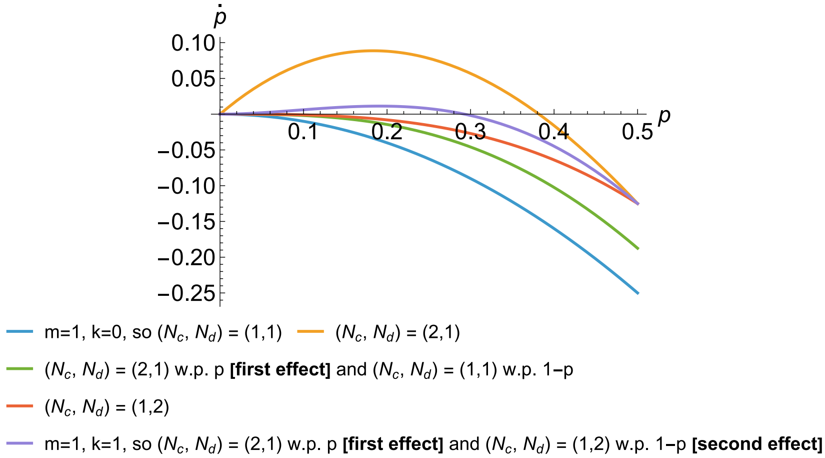


Fig. 3. Time derivative of p as a function of p. Consider $l, g < 1$. For $m = 1$, $k = 0$, universal defection is globally asymptotically stable. In contrast, when $m = 1$, $k = 1$, $p^* = \frac{1}{2}(2 - \sqrt{2}) \approx 0.293$ is almost globally stable. The effects of moving from the former to the latter, as discussed following Theorem 2, are also illustrated.

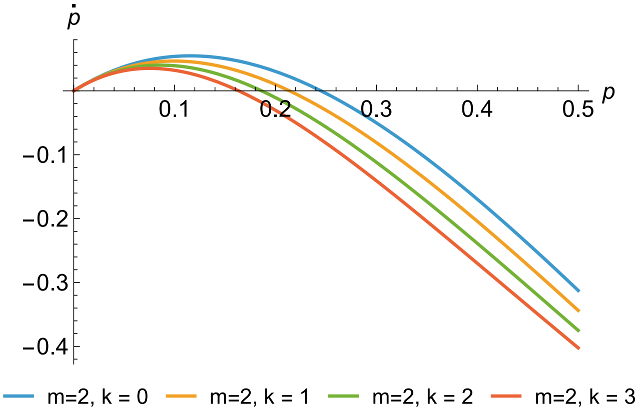


Fig. 4. Time derivative of p as a function of p. For $m = 2$, $l < \frac{1}{m-1}$, Theorem 3 shows that universal defection is unstable. For $k = 0, 1, 2, 3$, $l < \frac{1}{k+1}$, $g > k + 1$, we observe almost globally stable stationary states at $p^* \approx 0.245, 0.214, 0.186, 0.164$ respectively. Under these parameters, for C to be chosen by an entrant, at least one element of her C-sample must obtain payoff 1 and every element of her D-sample must obtain payoff 0.

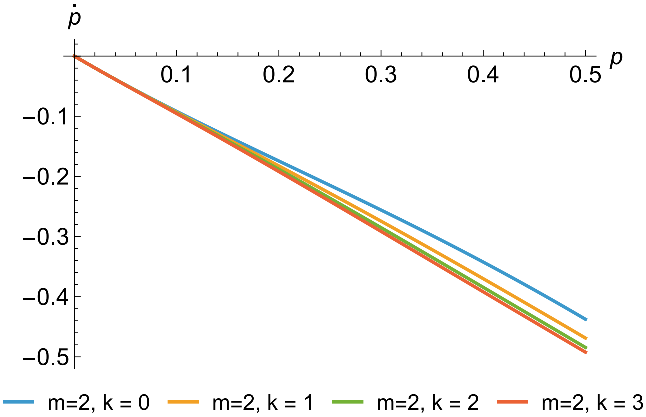


Fig. 5. Time derivative of p as a function of p. For $m = 2$, $l \geq \frac{1}{m-1}$, Theorem 3 shows that universal defection is asymptotically stable. Here we illustrate $k = 0, 1, 2, 3$, $l \geq k + 1$, $g > k + 1$. Under these parameters, for C to be chosen by an entrant, every element of her C-sample must obtain payoff 1 and every element of her D-sample must obtain payoff 0.

Table 2

Stable stationary states, p^*, of the fixed sample size dynamics for $g = l = 0.2$ and selected values of m, k. The rows for $m = 0$ and $m = 1$ illustrate Theorems 1 and 2 respectively. Further observe that an interior stationary state exists for $m = 0$, $k \geq 9$. Rows for $m \geq 2$ illustrate Theorem 3, with $m = 5$ and $m = 6$ illustrating the transition from $l < \frac{1}{m-1}$ to $l \geq \frac{1}{m-1}$. In addition, for large enough m, interior stationary states cease to exist. This is also the case for large k when $l \geq \frac{1}{m-1}$. In contrast, when $l < \frac{1}{m-1}$, interior stationary states continue to exist. Although they decrease towards zero as k gets large, this convergence is quite slow. For example, $m = 2$ and $k = 30$ still gives a cooperation share of 15.5%.

$m \backslash k$	0	1	2	3	4	9	30
0	–	–	0	0	0	0, 0.261	0, 0.181
1	0	0.293	0.316	0.257	0.371	0.284	0.163
2	0.282	0.246	0.362	0.287	0.288	0.237	0.155
3	0.332	0.269	0.309	0.249	0.296	0.209	0.129
4	0.356	0.280	0.289	0.275	0.228	0.195	0.105
5	0.372	0.272	0.220	0.239	0.230	0.195	0.065
6	0	0, 0.255	0	0	0, 0.207	0, 0.146	0
7	0	0	0, 0.224	0, 0.177	0, 0.162	0, 0.109	0
11	0, 0.216	0, 0.150	0, 0.103	0, 0.084	0	0	0
12	0	0	0	0	0	0	0

Table 3

Stable stationary states, p^*, of the threshold sample size dynamics for $g = l = 0.2$ and $m = 1$ to $m = 10$. The row for $m = 1$ illustrates Theorem 4. Rows for $m \geq 2$ illustrate Theorem 5, with $m = 5$ and $m = 6$ illustrating the transition from $l < \frac{1}{m-1}$ to $l \geq \frac{1}{m-1}$. In addition, for large enough m, interior stationary states cease to exist.

m	p^*	m	p^*
1	0	6	0, 0.158
2	0.308	7	0, 0.125
3	0.272	8	0, 0.083
4	0.242	9	0, 0.048
5	0.230	10	0

Theorem 4. *If $m = 1$, then $p = 0$ is asymptotically stable.*

It is instructive to consider the reason that this result differs from the fixed sample size result for $m = 1$ [Theorem 2]. Recall that the fixed sample size result depends on the rare occurrence of samples with $N_c > 1$. This event occurred with probability of order $\Theta(p)$. Now, consider threshold sample selection with $m = 1$. As soon as there is at least one sample of C and D, the process stops. Therefore, to obtain $N_c > 1$ we require the first two draws to be C. This occurs with probability p^2, which turns out to be too rare to destabilize universal defection.

Next, consider $m \geq 2$, so that sampling continues until there are at least two C and two D in the tested sample. Similar to the case of a fixed sample size [Theorem 3], it turns out that instability or stability of universal defection depends on whether a single C sample giving payoff 1 is enough for an entrant to choose C when all of the D samples give payoff 0.

Theorem 5. *Consider $m \geq 2$.*

- *If $l < \frac{1}{m-1}$, then $p = 0$ is unstable.*
- *If $l \geq \frac{1}{m-1}$, then $p = 0$ is asymptotically stable.*

Table 3 reports stable stationary states for $g = l = 0.2$ across selected values of m, computed numerically using Wolfram Research, Inc. (2026). The selected parameter values illustrate the different cases covered by Theorems 4 and 5. The table shows that, for $l \geq 1/(m-1)$, universal defection can coexist with an interior stable state, but this interior state eventually disappears as m increases.

4. Conclusion

This paper has introduced a broad class of decision rules that bridges procedural rationality and imitation, studied population dynamics under these rules for prisoner's dilemmas, and analyzed necessary and sufficient conditions for instability of universal defection. Results show that the stability of universal defection depends on the procedure by which strategies are sampled, the amount of direct information (procedural rationality) and indirect information (imitation) used by agents, and the game's payoff parameters.

The overall message of our results is the importance of testing strategies twice or more. As long as strategies are guaranteed to be tested twice or more and the potential loss risked by cooperation (l) is not too high, cooperation will proliferate when rare. This holds regardless of whether sample sizes are fixed or sampling is sequential until each strategy has been sampled some required minimum number of times.

In contrast, if the testing process is one of pure imitation, cooperation will never proliferate when rare. In the middle is the case where each strategy is guaranteed to be tested once. This typically leads to stability of universal defection, with one exception: if payoff incentives to defect (g) are weak, the damage from defection (l) is small and sample size is fixed, then cooperation proliferates when rare.

The "twice is enough" result can be read as a positive result for cooperation. Pure imitation leads to the stability of universal defection. Nevertheless, it does not take much procedural rationality to overturn this result. Guaranteeing that strategies are sampled at least twice is enough, regardless of how much of a factor imitation is in decision making.

Naturally, these are qualitative results. As we allow the amount of information used by agents to get large, either by guaranteeing through procedural rationality that strategies are sampled many times (large m) or by making it probable through imitation that they are sampled many times (large k), it becomes more and more probable that decision makers come to the conclusion supported by rational expectations: that defection leads to higher payoffs. This would lead to stable equilibria being close to universal defection, even if universal defection itself is unstable. Thus, our results are quantitatively important only as long as agents do not use large amounts of information.

The last point hints at another area of possible qualitative importance. Even when a society passes through a period of high information decision making, perhaps due to heightened alertness, worry or concern, as long as strategies are tested at least twice, some level of cooperation, no matter how small, will persist in the population. This opens the door to the possibility that when a calmer, low information decision making paradigm returns, there is some level of cooperative behavior present in the population that can then transition to a higher equilibrium value under the new information paradigm.

Agents in many economic and social settings mix direct testing with social observation. Overall, our analysis shows how the way information is generated and shared can determine whether defection persists or cooperation emerges. This suggests that interventions which alter learning procedures, not just incentives, may shift long-run outcomes. Applying the framework to games other than prisoner's dilemmas is a natural next step.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Jonathan Newton reports financial support was provided by JSPS. Srinivas Arigapudi reports financial support was provided by ANRF. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Proofs — fixed sample size

A.1. Simple random sampling $m = 0$, $k \geq 1$

Proof of Theorem 1.

Case $k = 1$.

Either the sample is entirely C, in which case the agent chooses C, or entirely D, in which case the agent chooses D. The former occurs with probability p, so we obtain $\dot{p} = p - p = 0$ everywhere. That is, every $p \in [0, 1]$ is a stationary state that is not asymptotically stable.

Case $k \geq 2$.

For the C-mean to exceed the D-mean, it is necessary that one of two mutually exclusive situations occur:

(I) All k draws are C. This occurs with probability p^k.

(II) Mixed sample ($1 \leq K_c \leq k-1$) and

  (i) at least one C receives payoff 1, and

  (ii) at least one D receives payoff 0.

This gives (numbering as above),

$$\dot{p} \;\leq\; \underbrace{\overbrace{p^k}^{(I)}}_{=O(p^k)} + \underbrace{\overbrace{\sum_{k_c=1}^{k-1}\binom{k}{k_c} p^{k_c}(1-p)^{k-k_c}}^{(II)}}_{=O(p^{k_c})} \underbrace{\overbrace{\left(1-(1-p)^{k_c}\right)}^{(i)}}_{=k_c p + O(p^2)} \underbrace{\overbrace{\left(1-p^{k-k_c}\right)}^{(ii)}}_{=O(1)} - p.$$

As $k \geq 2$, $p^k = O(p^2)$. Furthermore, $k_c \geq 1$ inside the sum, so the largest term is $O(p^2)$. Thus, for sufficiently small p,

$$\dot{p} \leq O(p^2) - p < 0,$$

and $p = 0$ is asymptotically stable. $\square$

A.2. Stratified sample size $m = 1$

Proof of Theorem 2.

Case $k = 0$

For the C-mean to exceed the D-mean it is necessary that C receives payoff 1 and D receives 0. That is,

$$\dot{p} \leq p(1-p) \; - \; p$$

which is < 0 for $p > 0$. Therefore, $p = 0$ is asymptotically stable.

Case $k \geq 1,\ g < k,\ l < 1$.

For the C-mean to exceed the D-mean it is sufficient that one of the following occur:

(I) $K_c = 0$. C receives payoff 1. All D receive 0.
Occurs with probability $(1-p)^k p(1-p)^{1+k} = p(1-p)^{1+2k}$.
(II) $K_c = 0$. C receives payoff 1. Exactly k of the D receive 0.
Occurs with probability $(1-p)^k p(k+1)p(1-p)^k = (k+1)p^2(1-p)^{2k}$.
(III) $K_c = 1$. At least one C receives payoff 1. All D receive 0.
Occurs w.p. $kp(1-p)^{k-1}(1-(1-p)^2)(1-p)^k = kp(1-p)^{2k-1}(2p-p^2)$.

Therefore,

$$\begin{aligned}
\dot{p} &\geq \overbrace{p(1-p)^{1+2k}}^{(I)} + \overbrace{(k+1)p^2(1-p)^{2k}}^{(II)} + \overbrace{kp(1-p)^{2k-1}(2p-p^2)}^{(III)} - p \\
&= p - (1+2k)p^2 + (k+1)p^2 + 2kp^2 + O(p^3) - p \\
&= kp^2 + O(p^3).
\end{aligned}$$

Thus, for sufficiently small p,

$$\dot{p} > 0,$$

and $p = 0$ is unstable.

Case $k \geq 1,\ g < k,\ l \geq 1$.

For the C-mean to exceed the D-mean it is necessary that one of the following occur:

(I) $K_c = 0$. C receives payoff 1. At least one D receives 0.
(II) $K_c \geq 1$. C receives payoff 1 at least twice.

Therefore,

$$\begin{aligned}
\dot{p} &\leq \overbrace{(1-p)^k p(1-p^{k+1})}^{(I)} \\
&\quad + \overbrace{\sum_{k_c=1}^{k} \underbrace{\binom{k}{k_c} p^{k_c}(1-p)^{k-k_c}}_{=O(p^{k_c})} \sum_{\kappa=2}^{m+k_c} \underbrace{\binom{m+k_c}{\kappa} p^{\kappa}(1-p)^{m+k_c-\kappa}}_{=O(p^{\kappa})}}^{(II)} - p \\
&= (1 - kp + O(p^2))p(1-p^{k+1}) + O(p^3) - p \\
&= -kp^2 + O(p^3).
\end{aligned}$$

Thus, for sufficiently small p,

$$\dot{p} < 0,$$

and $p = 0$ is asymptotically stable.

Case $k \geq 1,\ g \geq k$.

For the C-mean to exceed the D-mean it is necessary that the following occur:

(I) At least one C receives payoff 1.
This occurs with probability $1 - (1-p)^{1+k_c}$.
(II) All D receive payoff 0.
This occurs with probability $(1-p)^{1+k-k_c}$

Therefore,

$$
\begin{aligned}
\dot{p} &\le \sum_{k_c=0}^{k} \underbrace{\binom{k}{k_c} p^{k_c}(1-p)^{k-k_c}}_{=O(p^{k_c})} \underbrace{\overbrace{\left(1-(1-p)^{1+k_c}\right)}^{(I)}}_{=O(p)} \underbrace{\overbrace{(1-p)^{1+k-k_c}}^{(II)}}_{=O(1)} - p \\
&= \underbrace{(1-p)^{1+2k}p}_{\text{term for } k_c=0} + \underbrace{kp(1-p)^{2k-1}(2p-p^2)}_{\text{term for } k_c=1} + \underbrace{O(p^3)}_{\substack{\text{remaining} \\ \text{terms in sum}}} - p \\
&= p-(1+2k)p^2+2kp^2+O(p^3) - p \\
&= -p^2+O(p^3).
\end{aligned}
$$

Thus, for sufficiently small p,

$$\dot{p} < 0,$$

and $p = 0$ is asymptotically stable. □

A.3. Stratified sample size $m \ge 2$

Proof of Theorem 3.

Case $l < \frac{1}{m-1}$.

For the C-mean to exceed the D-mean it is sufficient that all of the following occur:

(I) $K_c = 0$. This occurs with probability $(1-p)^k$.
(II) At least one C receives payoff 1. Given (I), this occurs with probability $1-(1-p)^m$.
(III) All D receive payoff 0. Given (I), this occurs with probability $(1-p)^{m+k}$.

This gives

$$\dot{p} \ge \underbrace{\overbrace{(1-p)^k}^{(I)}}_{=1+O(p)} \underbrace{\overbrace{(1-(1-p)^m)}^{(II)}}_{=mp+O(p^2)} \underbrace{\overbrace{(1-p)^{m+k}}^{(III)}}_{=1+O(p)} - p = (m-1)p + O(p^2).$$

Thus, for sufficiently small p,

$$\dot{p} > 0,$$

and $p = 0$ is unstable.

Case $l \ge \frac{1}{m-1}$.

For the C-mean to exceed the D-mean it is necessary that at least two C in the sample receive payoff 1. This gives

$$
\begin{aligned}
\dot{p} &\le \sum_{k_c=0}^{k} \underbrace{\binom{k}{k_c} p^{k_c}(1-p)^{k-k_c}}_{=O(p^{k_c})} \sum_{\kappa=2}^{m+k_c} \underbrace{\binom{m+k_c}{\kappa} p^{\kappa}(1-p)^{m+k_c-\kappa}}_{=O(p^{\kappa})} - p \\
&= O(p^2) - p.
\end{aligned}
$$

Thus, for sufficiently small p,

$$\dot{p} < 0,$$

and $p = 0$ is asymptotically stable. □

Appendix B. Proofs — sequential sampling to threshold

B.1. Supporting lemmas

It will be useful to define two events, $\mathcal{E}_c$ and $\mathcal{E}_d$, defined as the final draw of the sequence being C and D respectively. Hence,

$$
\begin{aligned}
\Pr[\mathcal{E}_c] &= \sum_{k=m}^{\infty} \binom{m+k-1}{m-1} p^m (1-p)^k, \\
\Pr[\mathcal{E}_d] &= \sum_{k=m}^{\infty} \binom{m+k-1}{m-1} p^k (1-p)^m.
\end{aligned}
$$

Lemma 1. $\Pr[\mathcal{E}_c] = 1 - O(p^m)$, $\Pr[\mathcal{E}_d] = O(p^m)$

Proof.

For the last draw to be D, it is necessary and sufficient that at least m of the first $2m-1$ draws are C. Therefore,

$$\Pr[\mathcal{E}_d] = \sum_{k=m}^{2m-1} \underbrace{\binom{2m-1}{k} p^k (1-p)^{2m-1-k}}_{=O(p^k)} = O(p^m).$$

It follows that $\Pr[\mathcal{E}_c] = 1 - \Pr[\mathcal{E}_d] = 1 - O(p^m)$. $\square$

Lemma 2.

$$\Pr[\bar{C} > 0 \mid \mathcal{E}_c] = \begin{cases} p, & \text{if } \; m = 1 \\ mp + O(p^2), & \text{if } \; m \geq 2, \; l < \frac{1}{m-1} \\ O(p^2), & \text{if } \; m \geq 2, \; l \geq \frac{1}{m-1} \end{cases}$$

Proof.

Note that, conditional on $\mathcal{E}_c$, there are exactly m draws in the C-sample.

Case $m = 1$

There is only one draw in the C-sample. This draw has payoff more than zero if and only if it is played against C, which happens with probability p.

Case $m \geq 2, \; l < \frac{1}{m-1}$

As the C sample has size m, a single draw with payoff 1 gives a sample mean of

$$\frac{1-(m-1)l}{m} \underbrace{>}_{\text{by } l<\frac{1}{m-1}} 0.$$

Thus, at least one draw with payoff 1 is necessary and sufficient for $\bar{C} > 0$. The probability of this is

$$1 - (1-p)^m = mp + O(p^2).$$

Case $m \geq 2, \; l \geq \frac{1}{m-1}$

As the C sample has size m, a single draw with payoff 1 gives a sample mean of

$$\frac{1-(m-1)l}{m} \underbrace{\leq}_{\text{by } l\geq\frac{1}{m-1}} 0.$$

Thus, at least two draws with payoff 1 are necessary for $\bar{C} > 0$. The probability of this is

$$1 - (1-p)^m - mp(1-p)^{m-1} = 1 - (1 - mp + O(p^2)) - (mp + O(p^2)) = O(p^2). \quad \square$$

Lemma 3. *For all* $\delta \in (0,1)$*,* $\Pr[\bar{D} \geq \delta \mid \mathcal{E}_c] = O(p^m)$.

Proof.

When the D sample has size k, let r_k be the minimum number of $1+g$ payoffs required to make $\bar{D} \geq \delta$. That is,

$$r_k := \left\lceil \frac{\delta k}{1+g} \right\rceil.$$

Define

$$H(x,p) := x \log\left(\frac{x}{p}\right) + (1-x)\log\left(\frac{1-x}{1-p}\right).$$

Then (see, e.g. Arratia and Gordon, 1989) we have that

$$\Pr[\bar{D} \geq \delta \mid k] = \Pr[Bin(k,p) \geq r_k] \leq e^{-k\,H\left(\frac{r_k}{k},p\right)}.$$

Note that

$$\frac{r_k}{k} \geq \frac{\delta}{1+g} \quad \text{for all } k \in \mathbb{N},$$

and $H(x,p)$ is increasing in x on $x \in (p,1)$. Therefore, for $p < \frac{\delta}{1+g}$, we have

$$H\left(\frac{r_k}{k},p\right) \geq H\left(\frac{\delta}{1+g},p\right) =: h,$$

giving the bound

$$\Pr[\bar{D} \geq \delta \mid k] \leq e^{-k\,h} \quad \text{for all } k \in \mathbb{N}.$$

Summing over k,

$$\begin{aligned}
&\Pr[\bar{D} \geq \delta \mid \mathcal{E}_c] = \frac{1}{\Pr[\mathcal{E}_c]} \sum_{k=m}^{\infty} \binom{m+k-1}{m-1} p^m (1-p)^k \Pr[\bar{D} \geq \delta \mid k] \\
&\leq \frac{p^m}{\Pr[\mathcal{E}_c]} \sum_{k=m}^{\infty} \binom{m+k-1}{m-1} (1-p)^k e^{-k\,h} \\
&= \frac{p^m}{\Pr[\mathcal{E}_c]} \sum_{k=m}^{\infty} \binom{m+k-1}{m-1} (\underbrace{(1-p)e^{-h}}_{=:w<1})^k \\
&= \frac{p^m}{(1-w)^m \Pr[\mathcal{E}_c]} \underbrace{\sum_{k=m}^{\infty} \binom{m+k-1}{m-1} (1-w)^m w^k}_{\substack{\text{Tail probability of } -\text{ve binomial} \\ \text{distribution hence } <1}} \\
&< \frac{p^m}{(1-w)^m \underbrace{\Pr[\mathcal{E}_c]}_{\substack{=1+O(p^m) \\ \text{by Lemma 1}}}} \\
&= \frac{p^m}{\underbrace{(1-(1-p)e^{-h})^m}_{\substack{\to 1 \text{ as } p\downarrow 0 \\ \text{because } (1-p)e^{-h}\to 0}} (1-O(p^m))} = O(p^m), \quad \square
\end{aligned}$$

B.2. Threshold $m = 1$

Proof of Theorem 4.

We bound the probability of event $\bar{C} \geq \bar{D}$ together with events $\mathcal{E}_c$ and $\mathcal{E}_d$ separately, then combine the bounds.

First, on $\mathcal{E}_c$,

$$\begin{aligned}
&\Pr[\{\bar{C} \geq \bar{D}\} \cap \mathcal{E}_c] = \sum_{k=1}^{\infty} (1-p)^k \; p \; \Pr[\bar{C} \geq \bar{D} | (1,k)] \\
&\leq \sum_{k=1}^{\infty} (1-p)^k \; p \; \underbrace{\Pr[\bar{C} \geq 0 | (1,k)]}_{=p} = \sum_{k=1}^{\infty} (1-p)^k \; p^2 = p - p^2.
\end{aligned}$$

Second, on $\mathcal{E}_d$,

$$\begin{aligned}
&\Pr[\{\bar{C} \geq \bar{D}\} \cap \mathcal{E}_d] = \sum_{k=1}^{\infty} (1-p)\; p^k \; \Pr[\bar{C} \geq \bar{D} | (k,1)] \\
&= \sum_{k=1}^{4} (1-p)\; p^k \; \underbrace{\Pr[\bar{C} \geq \bar{D} | (k,1)]}_{\leq (1-p)(1-(1-p)^k)} + \sum_{k=5}^{\infty} (1-p)\; p^k \; \underbrace{\Pr[\bar{C} \geq \bar{D} | (k,1)]}_{\leq 1} \\
&\leq \underbrace{\sum_{k=1}^{4} (1-p)^2\; p^k \; (1-(1-p)^k)}_{=p^2-p^4+O(p^5)} + \underbrace{\sum_{k=5}^{\infty} (1-p)\; p^k}_{=p^5} \\
&= p^2 - p^4 + O(p^5).
\end{aligned}$$

Combining these bounds,

$$\begin{aligned}
&\Pr[\bar{C} \geq \bar{D}] = \Pr[\{\bar{C} \geq \bar{D}\} \cap \mathcal{E}_c] + \Pr[\{\bar{C} \geq \bar{D}\} \cap \mathcal{E}_d] \\
&\leq \left(p - p^2\right) + \left(p^2 - p^4 + O(p^5)\right) = p - p^4 + O(p^5).
\end{aligned}$$

Therefore,

$$\dot{p} \leq p - p^4 + O(p^5) - p = -p^4 + O(p^5),$$

which is strictly negative for small p. $\square$

B.3. Threshold $m \geq 2$

Proof of Theorem 5.

Case $l < \frac{1}{m-1}$

Choose

$$\delta \in \left(0, \frac{1-(m-1)l}{m}\right).$$

Then,

$$\begin{aligned}\Pr[\bar{C} > \bar{D}] &\geq \Pr[\mathcal{E}_c]\Pr[\bar{C} > \bar{D} \mid \mathcal{E}_c] \\ &\geq \underbrace{\Pr[\mathcal{E}_c]}_{\substack{=1+O(p^m) \\ \text{by Lemma 1}}} \left(\underbrace{\Pr[\bar{C} \geq \delta \mid \mathcal{E}_c]}_{\substack{=mp+O(p^2)\text{ by the} \\ \text{argument of Lemma 2}}} - \underbrace{\Pr[\bar{D} \geq \delta \mid \mathcal{E}_c]}_{\substack{=O(p^m) \\ \text{by Lemma 3}}} \right) \\ &= mp + O(p^2).\end{aligned}$$

Therefore,

$$\dot{p} \geq mp + O(p^2) - p = (m-1)p + O(p^2),$$

which is strictly positive for small p.

Case $l \geq \frac{1}{m-1}$

$$\begin{aligned}\Pr[\bar{C} > \bar{D}] &\leq \Pr[\bar{C} > 0] \\ &= \underbrace{\Pr[\mathcal{E}_c]}_{\substack{=1+O(p^m) \\ \text{by Lemma 1}}} \underbrace{\Pr[\bar{C} > 0 \mid \mathcal{E}_c]}_{\substack{=O(p^2) \\ \text{by Lemma 2}}} + \underbrace{\Pr[\mathcal{E}_d]}_{\substack{=O(p^m) \\ \text{by Lemma 1}}} \Pr[\bar{C} > 0 \mid \mathcal{E}_d] \\ &= O(p^2).\end{aligned}$$

Therefore,

$$\dot{p} \leq O(p^2) - p,$$

which is strictly negative for small p. □

Data availability

No data was used for the research described in the article.

References

Alós-Ferrer, C., Schlag, K., 2009. Imitation and learning. In: Anand, P., Pattanaik, P., Puppe, C. (Eds.), Handbook of Rational and Social Choice. Oxford University Press.
Apesteguia, J., Huck, S., Oechssler, J., 2007. Imitation—theory and experimental evidence. J. Econom. Theory 136 (1), 217–235.
Arieli, I., Arigapudi, S., 2024. Private signals and fast product adoption under incomplete information. Games Econom. Behav. 147, 377–387.
Arigapudi, S., 2024. Evolutionary dynamics in bilingual games. J. Econom. Dynam. Control 165, 104898.
Arigapudi, S., Heller, Y., 2025. Stable mixing in hawk–dove games under best experienced payoff dynamics. Games Econom. Behav. 151, 148–161.
Arigapudi, S., Heller, Y., Milchtaich, I., 2021. Instability of defection in the prisoner's dilemma under best experienced payoff dynamics. J. Econom. Theory 197, 105174.
Arigapudi, S., Heller, Y., Schreiber, A., 2025. Heterogeneous noise and stable miscoordination. Am. Econ. J.: Microeconomics 17 (4), 1–32.
Arigapudi, S., Lahkar, R., 2024. Emergence of trust in the trust game under best experienced payoff dynamics. Econom. Lett. 236, 111620.
Arratia, R., Gordon, L., 1989. Tutorial on large deviations for the binomial distribution. Bull. Math. Biol. 51 (1), 125–131.
Berkemer, R., Starke, J., Kawamoto, A., 2023. Sampling dynamics applied to the traveler's dilemma reveals non-Nash behavior. Games Econom. Behav. 142, 339–356.
Bowles, S., Gintis, H., 2011. A Cooperative Species: Human Reciprocity and Its Evolution. Princeton University Press, Princeton, NJ, USA.
Boyd, R., Gintis, H., Bowles, S., 2010. Coordinated punishment of defectors sustains cooperation and can proliferate when rare. Science 328 (5978), 617–620.
Cárdenas, J.C., Mantilla, C., Sethi, R., 2015. Stable sampling equilibrium in common pool resource games. Games 6 (3), 299–317.
Danenberg, T., Spiegler, R., 2025. A representative-sampling model of stochastic choice. J. Political Econ.: Microeconomics.
Duersch, P., Oechssler, J., Schipper, B.C., 2012. Unbeatable imitation. Games Econom. Behav. 76 (1), 88–96.
Fehr, E., Gächter, S., 2000. Cooperation and punishment in public goods experiments. Am. Econ. Rev. 90 (4), 980–994.
Glynatsi, N.E., McAvoy, A., Hilbe, C., 2024. Evolution of reciprocity with limited payoff memory. Proc. R. Soc. B 291 (2025), 20232493.
Hardin, G., 1968. The tragedy of the commons. Science 162 (3859), 1243–1248.
Heller, Y., Mohlin, E., 2018. Social learning and the shadow of the past. J. Econom. Theory 177, 426–460.
Hilbe, C., Nowak, M.A., Sigmund, K., 2013. Evolution of extortion in iterated prisoner's dilemma games. Proc. Natl. Acad. Sci. 110 (17), 6913–6918.
Lahkar, R., Mahanta, A., 2025. The best experienced payoff dynamic in the ultimatum minigame. Int. J. Game Theory 54 (1).
Mantilla, C., Sethi, R., Cárdenas, J.C., 2018. Efficiency and stability of sampling equilibrium in public goods games. J. Public Econ. Theory 22 (2), 355–370.

Nachbar, J.H., 1990. ''Evolutionary'' selection dynamics in games: Convergence and limit properties. Int. J. Game Theory 19 (1), 59–89.
Osborne, M.J., Rubinstein, A., 1998. Games with procedurally rational players. Am. Econ. Rev. 88, 834–847.
Osborne, M.J., Rubinstein, A., 2003. Sampling equilibrium, with an application to strategic voting. Games Econom. Behav. 45 (2), 434–441.
Oyama, D., Sandholm, W.H., Tercieux, O., 2015. Sampling best response dynamics and deterministic equilibrium selection. Theor. Econ. 10 (1), 243–281.
Salant, Y., Cherry, J., 2020. Statistical inference in games. Econometrica 88 (4), 1725–1752.
Sandholm, W.H., 2001. Almost global convergence to p-dominant equilibrium. Int. J. Game Theory 30 (1), 107–116.
Sandholm, W.H., 2005. Excess payoff dynamics and other well-behaved evolutionary dynamics. J. Econom. Theory 124 (2), 149–170.
Sandholm, W., 2010. Pairwise comparison dynamics and evolutionary foundation for Nash equilibrium. Games 1, 3–17.
Sandholm, W.H., 2010. Population games and evolutionary dynamics. In: Economic learning and social evolution, Cambridge, Mass. MIT Press.
Sandholm, W.H., Izquierdo, S.S., Izquierdo, L.R., 2019. Best experienced payoff dynamics and cooperation in the centipede game. Theor. Econ. 14 (4), 1347–1385.
Sandholm, W.H., Izquierdo, S.S., Izquierdo, L.R., 2020. Stability for best experienced payoff dynamics. J. Econom. Theory 185, 104957.
Sawa, R., Wu, J., 2023. Statistical inference in evolutionary dynamics. Games Econom. Behav. 137, 294–316.
Schipper, B.C., 2009. Imitators and optimizers in cournot oligopoly. J. Econom. Dynam. Control 33 (12), 1981–1990.
Schlag, K.H., 1998. Why imitate, and if so, how?: A boundedly rational approach to multi-armed bandits. J. Econom. Theory 78 (1), 130–156.
Schlag, K.H., 1999. Which one should I imitate? J. Math. Econom. 31 (4), 493–522.
Sethi, R., 2000. Stability of equilibria in games with procedurally rational players. Games Econom. Behav. 32 (1), 85–104.
Sethi, R., 2021. Stable sampling in repeated games. J. Econom. Theory 197, 105343.
Spiegler, R., 2006. Competition over agents with boundedly rational expectations. Theor. Econ. 1 (2), 207–231.
Wolfram Research, Inc., 2026. Mathematica, version 15.0.