

# Evolution without direction: evolutionary multistability in the iterated prisoner's dilemma

Matteo Smerlak\*

Max Planck Institute for Mathematics in the Sciences,  
Inselstrasse 22, Leipzig, D-04103, Germany

## Abstract

The iterated prisoner's dilemma (IPD) is a fertile model system to investigate the evolution of cooperation among self-interested agents, with applications ranging from microbiology to geopolitics. Conventional wisdom has it that selfish (or even extortionary) strategies give an edge in head-to-head IPD challenges, but that cooperative behavior is more stable at the population level. Here I show that the memory-one IPD—where phenotypes encode the stochastic response to just four possible joint plays—contains a phase where evolutionary dynamics is in fact multistable, with long periods of cooperation followed by rapid reversions to defection, and vice versa *ad infinitum*. This surprising, switch-like behavior shows that evolutionary dynamics does not always lead to evolutionarily stable states, even when phenotype space is small and easily explored.

## Introduction

In accord with Darwin's doctrine of the "survival of the fittest", evolution through natural selection is commonly viewed as a process directed towards adaptation. Thus, Wright's picture of the fitness landscape, where peaks correspond to successful genotypes and valleys to maladaptations, evokes an upwards-directed, fitness-maximizing process; Fisher's equally influential fundamental theorem of natural selection expresses the same basic principle that selection works to increase fitness. While this view of evolution as optimization does not imply that all traits must be adaptive (as famously noted by Gould and Lewontin, some fixed features appear to be byproducts of the evolution of other traits rather than genuine adaptations), it does suggest that irreversibility is a fundamental component of evolutionary change.

---

\*To whom correspondence should be addressed; E-mail: smerlak@mis.mpg.de

Several factors complicate this picture. For starters, genetic drift weakens the link between selective advantage and changes in allele frequency; thus, deleterious mutations can sometimes fix in small populations, as in e.g. Muller’s ratchet. Second, adaptation is not always a gradual process: it is well known that many species have undergone long periods of evolutionary stasis followed by short bursts of rapid adaptation. Moreover, the outcomes (and not just the times) of these bursts appear to be largely contingent. As Gould famously noted, it may well be that, should some supernatural being choose to “replay the tape of life”, the face of planet Earth would look completely different. The extent to which these contingencies are extrinsic (e.g. meteorites) or intrinsic (e.g. self-organized) has been a topic of lively discussion.

Third, and perhaps most importantly, the widespread occurrence of frequency-dependent selection—the fact that reproductive success can depend on traits other than one’s own—significantly muddles the picture of evolution as optimization. In particular, the replicator equation of evolutionary game theory does not necessarily lead to stable equilibrium points; simple examples, such as the rock-paper-scissors game, already give rise to periodic cycles. Nevertheless, evolutionary game theory has mostly focused on defining, identifying and characterizing evolutionarily stable states.

To address some of these towering questions, it can be useful to focus on simple systems with clearly interpretable phenotypes. The memory-one iterated prisoner’s dilemma (IPD<sub>1</sub>) provides one such system. In this game, two players X and Y will cooperate (C) or defect (D) with a probability that depends on the outcome of their previous interaction. If X and Y cooperate, they both earn a reward ( $R = 3$ ); if X cooperates and Y defects, then X is a ‘sucker’ ( $S = 0$ ) and Y falls to temptation ( $T = 5$ ); and if X and Y both defect, they both get a punishment  $P = 1$ . This simple rule gives rise to a surprisingly rich four-dimensional phenotype space  $\Sigma_1 = [0, 1]^4$  (where the first component is the probability to cooperate after CC, the second component is the probability to cooperate after CD, etc.), comprising classical strategies such as Tit-For-Tat (TFT) and Win-Stay-Lose-Shift (WSLS). Remarkably,  $\Sigma_1$  also contains extortionary strategies which take advantage of any opponent—no matter which strategy they choose—, as well as generous strategies which guarantee that the opponent never does worse than the player themselves.

Identifying which kind of strategy is evolutionarily stable in IPD<sub>1</sub> is an important problem in itself, as it may help understand how cooperation evolves in more realistic conditions. For a long time it was believed that, being simple, nice and retaliatory, TFT epitomizes an evolutionarily stable strategy in the iterated prisoner’s dilemma, and *a fortiori* in IPD<sub>1</sub>. The breakthrough discovery of so-called “zero-determinant” extortionary strategies within  $\Sigma_1$  by Press and Dyson [**press’iterated’2012**] forced a reappraisal of this conclusion [**ball’physicists’2012**]. Perhaps reassuringly, several studies found that more generous strategies can in fact replace extortionists in large populations [**adami’evolutionary’2013**, **stewart’extortion’2013**]. Whether this scenario does indeed pan out, or defection takes root permanently, turns out depend on the mutation rate and the strength of selection [**stewart’extortion’2013**, **iliopoulos’critical’2010**]; in particular, Iliopoulos *et al.* identified a phase transi-

tion at a critical mutation rate  $\mu_c$  between two phases, one dominated by cooperation ( $\mu \leq \mu_c$ , such that opponents are predictable enough to warrant the risk of cooperation), and the other by defection ( $\mu \geq \mu_c$ , in which opponents are too unpredictable).

In this letter I show that which strategy is more evolutionary stable in the IPD<sub>1</sub>—which direction evolution chooses to go given the mutation rate, population size, etc.—is not always well-posed. I describe a phase of IPD<sub>1</sub>, characterized by small mutational effects, weak genetic drift and no spatial structure, in which evolutionary dynamics is *multistable*. This means that, instead of converging to a single attractor within  $\Sigma_1$ , evolutionary trajectories shift unpredictably between two long-lived but nevertheless transient states. The resulting dynamics resembles the burst (or ‘popcorn’) noise observed in certain electronic devices, and can be modelled mathematically by a stationary, continuous-time Markov process over few effective states with exponentially distributed sojourn times. This behavior cannot be explained within adaptive dynamics.

## Results

**The phase diagram of evolution in the IPD.** I performed long-running agent-based simulations similar to those described in [iliopoulos’critical’2010]. Each agent is characterized by a stochastic strategy  $p = (p_1, p_2, p_3, p_4) \in [0, 1]^4 \equiv \Sigma_1$ , where  $p_1$  (resp.  $p_2, p_3, p_4$ ) is the probability to cooperate with the opponent after the joint play CC (resp. CD, DC, DD). At each generation, each agent plays a match of the infinitely iterated prisoner’s dilemma with  $T$  randomly chosen opponent in a population of size  $N$ . After each match, the payoff of a player playing strategy  $p$  against strategy  $q$  is increased by an amount  $\pi(p, q)$  which can be computed as a  $4 \times 4$  determinant given [press’iterated’2012]. Once all agents have played their matches, all strategies are mutated according to  $p'_i = \min(\max(p_i + \sigma\epsilon, 0), 1)$ , where  $\sigma > 0$  and  $\epsilon$  is a standard normal variable. After that, a new population of identical size  $N$  is sampled, with a weight proportional to the payoff accrued in the previous generation, and the entire process is repeated. In this formulation,  $\sigma$  measures the typical size of mutations and  $T$  the strength of selection.

Iliopoulos *et al.* describe the evolution of strategies in IPD<sub>1</sub> that are either cooperative or defective, depending on evolutionary parameters such as the mutation rate. I confirmed the existence of such a phase transition, measured here in terms of the rate of cooperation (the equilibrium probability of CC or CD plays)  $\langle C \rangle_t$  averaged over players, opponents and evolutionary time as a function of  $\sigma$  and  $T$  (Fig. XXX). As noted in Ref. [iliopoulos’critical’2010], the transition from a cooperative phase under weak mutations to a defective phase under strong mutations has a natural interpretation: when future opponents can be reliably assumed to be similar to previous opponents, i.e. when the environment of players is sufficiently predictable, cooperation provides higher payoffs. But when mutational effects are large, there is no knowing how the next generation of opponents will play, and cooperation becomes unacceptably risky. In this

phase evolution favors defection, at the cost of lower fitness for all.

Considering the variance of the mean population cooperation rate  $\text{Var}_t(C)$  over time as a function of  $\mu$  and  $T$  reveals a richer picture.

**Multistability and reduced effective dynamics.** How can an evolutionary state be stable for long stretches of time, and yet eventually yield to a radically different invader population spawn by small mutations only?

**Early-warning signals of evolutionary shifts.**

## Discussion

A large part of evolutionary theory in general, and evolutionary game theory in particular, has focused primarily on predicting the direction of evolution—or at least the phenotypes that evolution would reliably favour if its tape was replayed multiple times. Among these, the “evolutionarily stable states” introduced by Maynard Smith and Price, and their generalizations, have historically played a prominent role: being by definition uninvadable, these states correspond to the endpoints of evolutionary trajectories with game-like frequency dependence, even when they do not maximize payoff or fitness. Viewed through this angle, the problem of the evolution of cooperation is to determine whether, or under which condition, cooperative behavior is an evolutionarily stable strategy. The results presented in this letter show that Darwinian evolution through frequency-dependent selection does not always respect this paradigm: the iterated prisoner’s dilemma—the simplest and best studied setting to study the evolution of cooperation—contains a phase with multistable evolutionary dynamics.

This switch-like behavior is consistent with the observation that rapid, major evolutionary change can occur without any exogenous driver. Various explanations for such kind of “punctuated equilibrium” have been proposed. One, due to Sneppen and collaborators, relies on a simple but abstract and somewhat *ad hoc* model of self-organized critical evolutionary dynamics in a one-dimensional fitness space [sneppen’evolution’1995]. Another relates punctuations to the topology of genotype-to-phenotypes mappings: in very high-dimensional sequence spaces, it can take a long time for a population to diffuse along neutral networks until a new, fitter phenotype is found [fontana’continuity’1998, crutchfield’evolutionary’2002, bakhtin’evolution’2021]. A third explanation, sometimes called “quantum evolution”, depends on genetic drift to enable transient maladaptations before a new peak is found in the fitness landscape. In the multistable phase of IPD<sub>1</sub>, rapid evolutionary change arises due to yet another cause: XXX

Other aspects of the multistable behavior of IPD<sub>1</sub> are more surprising. While it is well-known that frequency-dependent selection is not constrained by Fisher’s fundamental theorem (i.e. that mean fitness is not generally monotonically increasing in evolutionary game theory), the sudden reductions in payoff observed here in populations with relatively large size are hardly expected. Of course, it is easy to construct

cooperative initial populations which are quickly invaded by defectors; what is remarkable here is that such reversions happen in populations that were selected for hundreds of thousands of generations.

Overall, our findings reveal a new mode of evolutionary dynamics in the iterated prisoner’s dilemma: one in which evolution has no attractor, no limit cycle and no preferred direction. In this phase, the evolutionary dynamics of cooperation may be compared with the reversible fluctuations of a physical system at thermal equilibrium—that is, to a process many biologists would characterize as antithetical to life.

## Methods

**Simulations.** Agent-based simulations were coded and run using the Agents.jl julia package [datseris’agents’2021].