

Strategic evolution in extensive form games

June 5, 2025

Abstract

Evolutionary game theory is often agnostic about its interpretation. One set of equations can represent biological evolution, cultural evolution, or individual learning. We argue that this multiple interpretation cannot hold in games with a temporal structure (known as sequential or extensive-form games). We introduce a new dynamic which better represents evolution and learning in extensive-form games, and we illustrate some differences. Our model also provides a condition under which evolutionary game theory will coincide with classical game theory.

1 Introduction

Evolutionary Game Theory represents one of several foundational approaches to game theory. It comprises a set of models that represent how low-rationality or no-rationality players will learn or evolve to play a game. The replicator dynamic is one such model.

The replicator dynamic has two interpretations: a biological interpretation and a learning interpretation. Under the biological interpretation, players are fixed phenotypes in a population. The payoff in the game represent the biological fitness of those phenotypes. Strategies change over time because the phenotypes are subject to natural selection.

Under the learning interpretation, individual players have strategies which change over time. This change is governed by a process of cultural evolution by imitation [11], a version of reinforcement learning [1], or a process of conscious deliberation [14].

The power of the replicator dynamic lies in its ability to represent all of the above interpretations with a single set of mathematical equations. Researchers can discuss issues like the evolution of cooperation, fairness, meaning, or division of labor without committing themselves to a particular interpretation. One model generates both biological-, cultural-, and learning-based explanations.¹

¹The replicator dynamic provides a clear illustration of how a model is made up of both a mathematical structure and an interpretation [18].

For example, Skyrms shows that correlation in the prisoner’s dilemma is sufficient to lead to the evolution of cooperation [13]. This claim is valid under all three interpretations of the dynamic, and so he has provided a framework for explanation of the biological evolution, the cultural evolution, and (in some limited contexts) the learning of cooperation in interactions well modeled as prisoner’s dilemma games.

Usually, in evolutionary game theory, games are considered in their “normal form.” This represents players’ choices as taking place simultaneously or, more precisely, in ignorance of what other players have chosen. Many of the canonical games, like the Prisoner’s Dilemma, Stag Hunt, and Nash demand games, are simultaneous move games.

However, there is a broader class of games where players make decisions with at least some information about the moves other players have taken in front of them. These games are called sequential or extensive-form games. Canonical recreational games like chess, go, and poker are of this form. So too are other important games in game theory, like the repeated Prisoner’s Dilemma, ultimatum bargaining game, or Lewis signaling games.

One can transform an extensive-form game into normal form by representing a player’s strategy for every possible contingency. It specifies what a player will do in response to all possible moves by an opponent.

In simpler games, under the biological interpretation, this conversion makes sense. If a player’s strategy is encoded in their genes, the information about all their contingency plans is already there and transmitted to the next generation regardless of whether the contingency occurs.

It is less clear that using the normal form is appropriate in the context of learning or imitation. Consider chess. A normal form strategy would have to specify a move for every possible board position, something that is computationally completely intractable. When a player is contemplating a move in chess, or imitating the play of another player, they do not learn or imitate the entire contingency plan. They adapt only part of their contingency plan that they are considering or that they observe.

For example, chess players can consider puzzles — board positions that they have never seen — and try out local strategies in those isolated contexts. In the context of the full game tree of chess, these agents are effectively adapting a part of their contingency plan at just that local decision (or set of local decisions). Different agents might make this adaptation at different rates (or by different amounts). A purely imitation-driven agent might only change their local decision as often as they actually encounter it in real play. A deliberative agent might deliberate about what they would do in an unobserved contingency. These possibilities are excluded by the replicator dynamic.

In this paper, we present a generalization of the replicator dynamic which is more suited for the learning interpretation in extensive-form games. We show that this dynamic is relevantly different from the replicator dynamic. For one version of the dynamic, the play converges to the subgame perfect Nash equilibrium (something the replicator dynamic is known not to do). Another version of the dynamic is equivalent to the replicator dynamic for some classes of games. We therefore suggest that this generalization is a more appropriate model when

considering extensive form games from a learning perspective. This generalization illustrates that the tri-fold interpretation of the replicator dynamic cannot persist in extensive form games.

2 Evolutionary Game Theory

A common assumption in much of classical game theory is that the players of a game are rational in some way and reach an equilibrium through this rationality. For example, in the Prisoner’s Dilemma, there is a unique Nash equilibrium in which both players choose Defect. Under classical game theory, the players would arrive at this through some reasoning process (e.g. elimination of dominated strategies).

Evolutionary game theory does not make this assumption; instead, the players initially play some arbitrary (mixed) strategy, and then revise this strategy by increasing the proportions of pure strategies that performed well while decreasing the proportions of pure strategies that performed poorly.

Evolutionary game theory is a broad class of mathematical models and attendant interpretations of those models. In these models, the players can be interpreted either as individual players of the game or as populations. A mixed strategy can be interpreted as an individual player’s conjecture over their own strategy, a player’s conjecture of an opponent’s strategy, an objective probability distribution over their own strategy, or the state of a population of players who are themselves adopting either a pure or mixed strategy. The dynamic process by which those strategies change can be interpreted as a biological evolutionary process, a cultural evolutionary process governed by imitation, or an individual learning process.

We will focus attention on one particular mathematical model which can be interpreted in many of these different ways: the replicator dynamic. Originally the replicator dynamic was used as a model of biological evolution via natural selection— strategies grow or die based on their relative Darwinian fitness to the current population [15]. However, the replicator dynamic can be interpreted differently: it can represent a learning strategy which “seeks the good” in the language of Skyrms [14].

Consider a two-player normal-form game with pure strategies for player 1 $S_1 = \{a_1, a_2, \dots, a_n\}$ and pure strategies for player 2 $S_2 = \{b_1, b_2, \dots, b_m\}$. Let p_i and q_j represent the proportions of strategies a_i and b_j , respectively. Then the continuous-time replicator dynamic is given as the following set of differential equations:

$$\dot{p}_i = p_i(f_i^1 - \phi_1), \quad \dot{q}_j = q_j(f_j^2 - \phi_2),$$

where $f_i^1 = \sum_j q_j \pi_1(a_i, b_j)$ is the expected payoff or *fitness* of pure strategy a_i (and similarly for f_j^2), and $\phi_k = \sum_i \sum_j p_i q_j \pi_k(a_i, b_j)$ is the average fitness (or expected payoff under the current mixed strategies) of Player k .

A slightly more compact notation which will be adapted in later parts of the paper is the following:

$$\dot{p}_i = p_i(e_i - p) \cdot Aq, \quad \dot{q}_j = q_j(f_j - q) \cdot Bp,$$

only. But this is an innocuous assumption within the confines of a mathematical analysis [17]

This position was controversial. Perhaps most notably, are the introduction of equilibrium refinements, like subgame perfection, which are only defined using the extensive form game [6].

As an example consider the sequential game pictured in Figure 1 (referred to as the Mini-Ultimatum Game). Player 1 makes an offer to Player 2 for how to divide \$10. Player 1 is restricted to either making a “fair” offer where they split it equally or making an “unfair” offer where Player 1 keeps \$8 and Player 2 gets \$2. Player 2 can accept the offer or refuse, in which case both players get nothing.

In this game, Player 2 has two information sets, one for each potential offer from player one with two actions in each (A_1, R_1 in set 1 and A_2, R_2 in set 2). To represent this game in the normal form, we must specify what player 2’s contingency plan is at each of these information sets. This results in a total of four different normal-form strategies: $A_1A_2, A_1R_2, R_1A_2, R_1R_2$.

In its normal form pictured in Figure 2, the game has two (infinite) sets of Nash equilibria. In one set, Player 1 plays U and Player 2 plays any mixed-strategy combination of A_1A_2 and A_1R_2 , and in the other set, Player 1 plays F and Player 2 plays mixed-strategy combinations of A_1A_2 and R_1A_2 , but with the added constraint that $\Pr[A_1A_2] \leq 5/8$.

However, there is a unique subgame-perfect equilibrium at the point (U, A_1A_2) . The reasoning is that Player 2 would always choose A in any of his decisions, as it is the final decision and A always leads to a higher payoff than R . Knowing this, Player 1 would thus always choose U . More generally, this backward induction reasoning process starts by determining what players should do at the terminal decisions of a game, and then ‘truncates’ the game tree by removing all other actions from those decisions (effectively removing one layer of the game tree). It’s clear that this process heavily relies on additional structure provided by the game tree, and cannot be achieved using only the normal form.

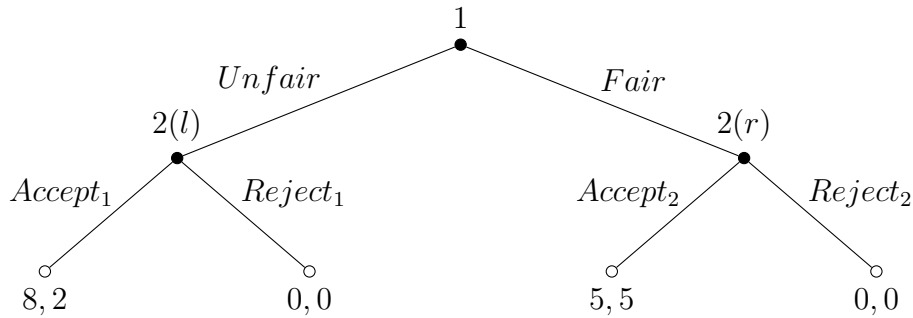


Figure 1: Mini-ultimatum game

Simon Huttegger has recently given a very different argument for the non-equivalence between extensive and normal form representations [9, 8]. He shows how a concept of stability to small perturbations (known as structural stability) should be interpreted differently in the context of normal and extensive form. His illustrations use the replicator dynamic, but

	A_1A_2	A_1R_2	R_1A_2	R_1R_2
U	8, 2	8, 2	0, 0	0, 0
F	5, 5	0, 0	5, 5	0, 0

Figure 2: The normal form of the mini-ultimatum game

would be equally valid in our novel dynamic. We, therefore, view his arguments as mutually compatible with ours.

In contrast to classical game theory, most of the existing literature on evolutionary dynamic analyze sequential games in their normal form [3]. The existing method of analysis for sequential games is to first convert it into a normal form, and then analyze the normal form with the replicator dynamic. The strategies specified by the replicator dynamic are the normal form strategies.

From the biological interpretation, this is perfectly reasonable. One would assume that genes encode normal-form strategies. As a result, those entire contingency plans are inherited. However, from a learning or imitation interpretation the players might adapt only parts of their normal-form strategy—i.e., actions at certain information sets—while keeping other parts the same. This is not viable under the current treatment of sequential games in evolutionary models.³

One consequence of this way of analyzing games is that there is no guarantee that the replicator dynamic will converge to a sub-game Nash equilibrium in a game. Consider a simple example, called the chain store game (pictured in Figures 3 and 4). Player 2 represents a store that is in an existing market that player 1 is contemplating entering. Player 1 has to decide whether to enter a market. If player 1 does not enter the market, then player 2 continues to monopolize the market. If player 1 enters, player 2 must decide whether to fight or acquiesce. If player 2 fights, they are both harmed. If player 2 acquiesces they split the market.

In this game there are two sets of Nash equilibria. One, the subgame perfect equilibrium, is where player 1 enters and player 2 acquiesces. However, there is a set of non-subgame perfect equilibria where player 1 stays out and player 2 commits to fighting with a sufficiently high probability.

The replicator dynamic can converge to the set of Nash equilibria which are not subgame perfect [3]. This occurs when the initial state of the population involve sufficiently high probability of player 2 fighting. In that case, player 1 evolves toward staying out. Although there is still selective pressure toward acquiescing the magnitude of this decreases the more that player 1 stays out.

³In games where a player takes one action and then potentially takes a later action, the normal form can be simplified slightly. If a poker player is resolved to never engage in a check-raise, then they do not need to form a contingency plan for what to do after they have check-raised. As an illustration consider the Centipede game (length 4) shown in Figure 7. In this game the “reduced normal form” treats certain normal-form strategies like D_1D_2 and D_1A_2 — which specify the same path within the game tree — as a single strategy D_1 .

concept of an “agent normal form”. [10, 12]. However, Selten mainly uses this agent normal form to prove the existence of (extensive-form) trembling hand perfect equilibria. Our model runs the replicator dynamic over this agent normal form, and allows for additional adjustments to the rates of evolution of information sets (which we will discuss more thoroughly in Section 5).

A different evolutionary model that uses the replicator dynamic for extensive form games is given by Gatti et al [5]. There, the authors implement a replicator dynamic over the *sequence form* of a game [16], which considers as strategies all sequences (including partial or non-terminal ones) of moves for a player. This model is significantly different from our and we will not delve into the differences.

4 New Model

There is no concern about representing a strategy in the normal form—this is a perfectly coherent way to understand strategies. Our concern is that modeling evolution on the normal form strategies is restrictive; it conceives of evolution and learning in one way that might not be fully general. It requires that normal form strategies evolve, are learned, or are imitated in totality.

To illustrate our concern imagine that you are player 2 in the mini ultimatum game pictured in Figures 1 and 2. You currently employ the strategy R_1R_2 , you always reject. You observe another player who was offered the fair offer and accepts. You see they have done better than you, and you resolve to imitate their strategy.

In the normal form, this imitation would have you imitate not only their acceptance of the fair offer but also what they *would have* done had they received the unfair offer—something you never observed.

An alternate notion of a strategy that better suits our purposes here is the behavioral strategy, which directly assigns a probability of what the player would do at each information set. More formally, a behavioral strategy b_n of player n assigns a probability distribution b_n^u on the set of actions C_u for every information set $u \in \mathcal{U}_n$. The space of behavioral strategies for a dynamic game is a subspace of the full mixed strategy space that is a product of simplices at each information set — in other words, every behavior strategy b naturally corresponds to a mixed strategy (the *canonical mixed strategy* of b), but there are many different mixed strategies that correspond to the same behavioral strategy through marginal probabilities.

An example that illustrates the difference between behavioral strategies in the mini ultimatum game. In the normal form of this game, Player 2 has four pure strategies: A_1A_2 , A_1R_2 , R_1A_2 , R_1R_2 . The mixed strategy space is thus a 3-simplex or tetrahedron (defined by $\sum_{i \in S_2} p_i = 1$). However, the set of behavioral strategies is only a square (the product of two 1-simplices), which lies inside the mixed strategy space.⁴ Every behavioral strategy

⁴Readers familiar with evolutionary game theory may be more used to seeing phase portraits that are triangular simplices, which represent the single probability space over three (typically) or more elements.

corresponds to a mixed strategy in the natural way — that is, the probability for each pure strategy is just the product of the probabilities of each individual component selected at their corresponding information sets. On the other hand, not every mixed strategy can be obtained this way through a behavioral strategy. As an example in this game, both mixed strategies $(1/4, 1/4, 1/4, 1/4)$ and $(1/2, 0, 0, 1/2)$ have *marginal* probabilities of $(1/2, 1/2)$ in both the left and right information sets, but only the former is realized by the behavioral strategy.

Since we want evolution to happen at the level of information sets, we will define a version of the replicator dynamic on the subspace of behavioral strategies instead. That is, the *induced replicator dynamic*⁵ for a dynamic game Γ is defined over the set of all information sets as the following set of equations:

$$\dot{p}_i^u = K^u p_i^u (\pi_n(p^u \setminus e_{u,i}) - \pi_n(p^u)).$$

Here, p^u denotes the mixed-strategy profile restricted to the information set u (i.e., only mixed strategy profiles that reach u are considered), and $p^u \setminus e_{u,i}$ denotes the mixed-strategy profile that is equal to p^u at all information sets other than u , and plays the pure local behavior strategy $e_{u,i}$ at u . In other words, $\pi_n(p^u)$ is Player n 's expected payoff starting from u , and $\pi_n(p^u \setminus e_{u,i})$ is Player n 's expected payoff upon playing $e_{u,i}$ at u .

There are two other notable differences between this new dynamic and the standard replicator dynamic: not all of our “strategies” now start at the beginning of the game, and not all of them will have actions that lead to a payoff node.

For the first difference, we introduce a weighting parameter for each information set K^u , which is a function (of the proportions of other strategies) that can be adjusted as a measure of the speed of learning or evolution at each information set. Intuitively, for information sets that lie after the start of the game, the speed of evolution should be scaled by the probability of reaching that point — this corresponds to a imitative or “low rationality” version of the model. On the other hand, if players know and can reason about the structure of the game, then the speed of evolution at all information sets should be the same — this is the “high rationality” version of the model. Our two possible weighting constants are thus the *realization probability* of the information set or 1 (which we will also refer to as the *full probability*).

Since we allow K^u to represent the realization probability, K^u is strictly a function from (some) p^u 's to $[0, 1]$. That is, the probability of reaching an information set u will depend on the probabilities of players who act before u and whose actions determine whether or not u will be reached. While K^u can vary in this way, we will suppress this in the notation for readability.

The space of behavioral strategies is a *product* of probability spaces, one for each information set. For simple games like the one below where each information set contains only two actions, the simplex looks like a hypercube (i.e., a product of many lines).

⁵The reason for this name is because the new dynamic is ‘induced’ by the standard replicator dynamic over the subspace of mixed strategies containing only *behavioral strategies*.

For the second difference, since the behavioral strategy gives probabilities at all information sets (including the ones after the one in question), we can just naturally define the payoffs of non-terminal nodes as the expected payoff with respect to these probabilities.

As an example of this new dynamic, consider the Mini-ultimatum game from 1 Player 2 has two different equations that govern the evolution of his decision (specifically, the time derivative of one of his pure local behavioral strategy) at each information set:

$$\begin{aligned}\dot{p}_{A_1}^l &= p_U p_{A_1}^l (\pi_2(A_1) - (p_{A_1}^l \pi_2(A_1) + p_{R_1}^l \pi_2(R_1))) = p_U p_{A_1}^l (2 - (2p_{A_1}^l + 0)) \\ \dot{p}_{A_2}^r &= p_F p_{A_2}^r (\pi_2(A_2) - (p_{A_2}^r \pi_2(A_2) + p_{R_2}^r \pi_2(R_2))) = p_U p_{A_2}^r (5 - 5p_{A_2}^r)\end{aligned}$$

5 Results

With a new model in hand, we can ask whether the induced replicator dynamics is different from the standard replicator dynamics. In one sense the answer is obviously “yes,” since they are defined over different objects (behavioral versus normal-form strategies). The differences are more significant than this. The two dynamics can, in some circumstances, lead to different predictions about the long-run play in some games. Whether they diverge depends on the game and the K^u -parameters. In this section, we explore some important similarities and differences.

Since the standard replicator dynamics gives the evolution of pure strategies in the normal-form game while the induced replicator dynamic gives the evolution of local behavioral strategies in the extensive-form game, we need to perform a transformation of variables from one to the other in order to compare the two models. In the results below, we compare the evolution of local behavioral strategies in the standard replicator dynamic and the two different versions of the induced replicator dynamic (i.e., using realized probabilities and full probabilities). Thus, when we talk about the models directly (e.g., the two models are equivalent, or differ by some factor), we are referring to the evolution of a local behavioral strategy S_u at information set u for Player i .

First we will consider *elementary, two-player, perfect information games*. These are extensive form games where Player 1 takes an action. One action will terminate the game immediately while the other action will lead to a response by Player 2. In that case, Player 2 can take one of two actions which also immediately terminates the game. (See Figures 5 and 6.) These games represent the simplest possible extensive form games that nonetheless have interesting game theoretic properties.

Although formally quite simple, elementary perfect information games nonetheless include several games of philosophical and scientific interest. The chain store game, trust game, and the centipede game of length two are elementary perfect information games and are used to study threats and cooperation.

The two dynamics will be equivalent when K^u is set to the “realized probability” of the information set. This means that for each information set K^u is set to be the probability that, given the strategy for the other players, the information set is reached. In elementary perfect information games, this means $K^1 = 1$ and $K^2 = p_R$

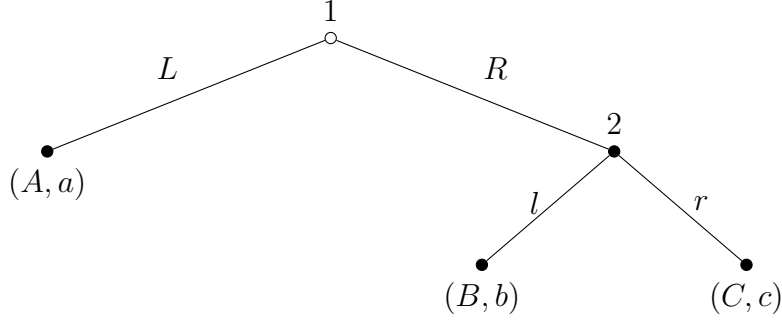


Figure 5: Elementary perfect information game in extensive form

	l	r
L	A, a	A, a
R	B, b	C, c

Figure 6: Elementary perfect information game in normal form

Lemma 1. For all elementary two-player perfect information games, the standard replicator dynamic is equivalent to the induced replicator dynamic using realized probabilities.

This indicates one similarity between the two dynamics. If we set K^u to be the probability that an information set is reached, then the two dynamics are equivalent and the new model has identical predictions to the classic replicator dynamic.

Next, we compare the two dynamics over two different classes of extensive-form games—the general (n -offer) Ultimatum game and the general (length- n) Centipede game (whose extensive forms are shown below).⁶

The n -offer Ultimatum game is like the Mini-Ultimatum game pictured in figure 1, except there are more potential offers from the first player. These offers can be any division of the good. Rejection always gives both players nothing.⁷

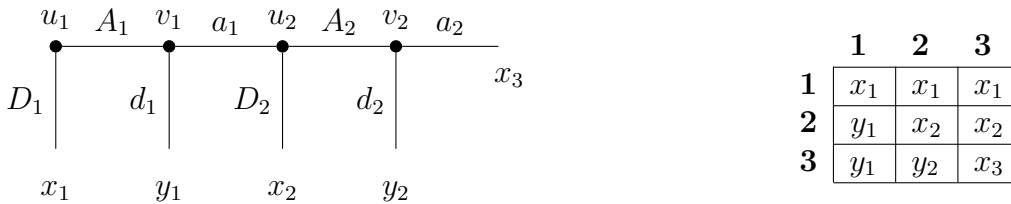


Figure 7: Centipede Game of length 4

The length- n centipede game describes a game where there is a pool of money. The players

⁶Note that Lemma 1 is actually just an instance of Theorem 3 below, since all self-realization probabilities are 1. Nonetheless, we present it separately as a simple demonstration of the larger result below.

⁷Strictly speaking our proof is for an even more general class of games. The proof covers any game where player 1 has n actions, after player 1 takes action i , player 2 has the choice of two actions one (“accept”) that pays s_i to player 2 and another (“reject”) that pays r to player 2 where $s_i > r$. r is constant across all actions by player 1, while s_i depends on the choice by player 1. Player 1 can have any payoffs.

alternate moves, and each player can choose between ending the game now and taking more than half of the money in the pool or passing to the other player. With each “pass” the amount of money grows. The game is set up so that if player i knows that player j will end the game on the next turn, player i does better by ending the game on this turn rather than passing. Whoever has the n^{th} (last) move also does better by ending rather than passing and ending the game. The length-4 centipede game is pictured in figure 7.

Both games are well-known in the economics and philosophy literatures as examples in which the theoretically rational outcome (subgame perfect equilibrium) does not match actual or desired human behavior. Because our model allows for different values of the weighting parameter—one of which leads to convergence to the SPE—we can think of it as a potential explanation for this observed phenomenon. That is, classical game theory assumes a very high degree of rationality of its agents, while in reality this degree is often lower, as in the standard replicator dynamic model.

The games also pit normatively desirable human qualities (like fairness and cooperation) against self interest. In the case of the ultimatum game there is a conflict for player 1 between making a “fair” offer and taking more for herself. In the centipede game, there is a conflict between cooperation (passing to increase the payoff to both players) and self-interest (taking now to increase one’s share). The centipede game can be thought of as a simplification of the finitely repeated Prisoner’s dilemma.

Structurally these games display the two different kinds of relations that information sets of the same player can have within a game. If two information sets of the same player lie on the same path in the game tree, they are *dependent*; if they do not lie on the same path, they are *independent*.

In the Ultimatum game, Player 2’s information sets are all *independent*—that is, any path in the game tree will encounter at most one such information set. In the Centipede game, both players’ information sets are *dependent*—that is, for any two information sets of the same player, there exists a path that contains both. As we will show below, this difference in structure within the game tree will impact the evolution of local behavioral strategies at the information sets under the induced replicator dynamic.

Theorem 2. For the n -offer Ultimatum game, the standard replicator dynamic and induced replicator dynamic using realized probabilities are equivalent.

In other words, when all information sets are independent, the induced replicator dynamic on the extensive-form game actually yields the same evolutionary process as the standard replicator dynamic on the normal-form game. Equivalently, evolution under the standard RD behaves *as if* local behavioral strategies evolved under the induced RD with realized probabilities.

Turning to the centipede game, strategies arising from dependent information sets are tricky to handle. The complication arises because a player’s choice at an information set determines whether their later information sets are reached. If player 1 chooses D at their first opportunity, then they will not reach their second information set. Thus, some game theorists choose to treat strategies that differ only in these unentered information sets as the same strategy.

- Full probabilities⁹
 - **Evolutionary:** An individual always updates their strategy every time the game is played.
 - **Rational:** An individual believes the game always reaches their information set (or rather, the individual can always reason about their strategy at this information set).
- External realized probabilities
 - **Evolutionary:** An individual updates their strategy when the game reaches their information set. All teammates along this individual’s path always continue the game towards my information set.
 - **Rational:** An individual believes all previous moves made by other individuals on that individuals “team” must send the game towards their information set.

Returning to the centipede game, we now can state the following theorem.

Theorem 3. For the centipede game of length- n , the standard replicator dynamic and induced replicator dynamic using realized probabilities differ by a factor of K_i^u , which denotes the self-realized probability — i.e., the part of the realized probability containing *only* Player i ’s local behavioral strategies — for node u .

Figure 10 shows the difference the centipede game of length 3 in the standard and induced replicator dynamics with realized probabilities.

An immediate consequence of theorem 3 is that, if we replace the realized probabilities with external realized probabilities in the induced RD, we would get evolution equivalent to the standard RD over the normal form. This provides one illustration of how one might justify the replicator dynamic in the context of learning or cultural evolution.

Corollary 1. For the n -offer Ultimatum game and the Centipede game of length n , the standard replicator dynamic and the induced replicator dynamic using external realized probabilities are equivalent.

This illustrates the first difference between the two dynamics. However, in terms of the resulting behavior the difference is not substantial. Our last result shows a more significant difference between the two dynamics. In particular, it illustrates how the induced replicator dynamics can underwrite classical game theory in a way that the standard replicator dynamic does not.

Theorem 4. For all generic PI games, the induced replicator dynamic using full probabilities has a unique globally asymptotically stable point, which corresponds to the unique subgame-perfect equilibrium of that game.

This illustrates how our dynamic both generalizes the replicator dynamic and provides a way that it can coincide with classical game theory. When the induced dynamic uses realized

⁹As a sidenote, this version of the new dynamic is equivalent to the standard replicator dynamic on Selton’s agent normal form

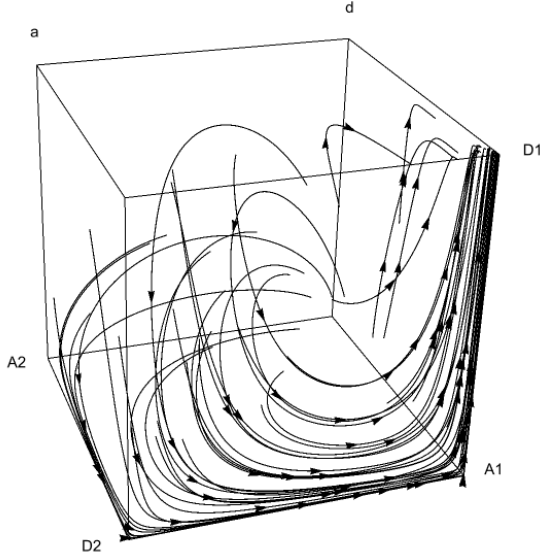


Figure 8: C3 game, standard RD

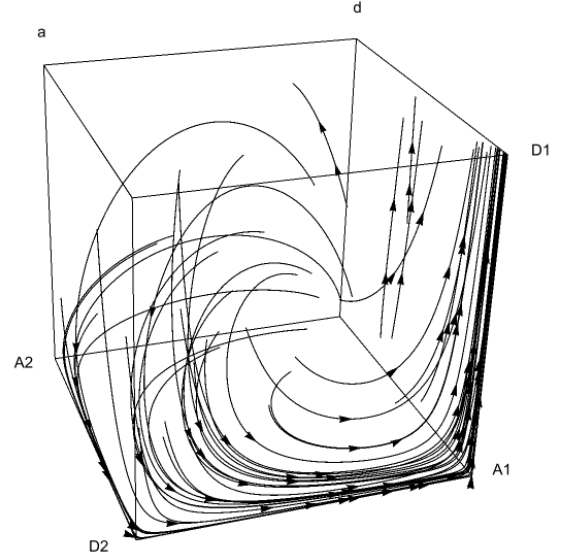


Figure 9: C3 game, induced RD

Figure 10: The centipede game of length 3 in the standard and induced replicator dynamics using realized probabilities. Although the dynamics are different, they have similar trajectories — in particular when p_D is close to 0 (corresponding to populations that are playing into the subgame) they are almost identical. As p_D increases (i.e., moving upwards in the simplex) the two dynamics become more different. As p_D nears 1 motion along the $A_2 - D_2$ direction slows down in the new dynamic (because the behavioral strategies are no longer playing into the subgame).

probabilities, in the chain store game (Figures 3 and 4) the dynamic is not guaranteed to converge to the subgame perfect equilibrium. However, using full probabilities the induced replicator dynamic will converge to the subgame perfect equilibrium.

The intuitive difference between the full probability and realized probability settings is that in the latter, strategies can evolve such that certain information sets are never reached. If an information set is never reached, there will be no payoff difference between strategies that only differ in that information set. Under the standard replicator dynamic, therefore, local action at these sets will no longer change.

For example, in the Mini-Ultimatum game, it’s possible for Player 1’s strategy to reach F entirely, which would mean that Player 2 never reaches the left subgame. This allows for convergence to NE that are not subgame perfect, such as $(F, R_1 A_2)$. However, using full probabilities prevents evolution from prematurely terminating in this fashion — even at $(F, R_1 A_2)$, Player 2’s strategy would continue to evolve until the left action becomes A_1 . The resulting process is essentially the same as backward induction, and leads to convergence at the SPE.

6 Conclusion

The standard way of modeling sequential games is to run the replicator dynamic over the set of normal-form strategies, which in many contexts (like the chess example in the introduction) is too coarse of an evolutionary process than what we’d like. In this paper, we introduce a more generalized way to use the replicator dynamic by running it over the set of local behavioral strategies or *actions* in the game tree. This way, we capture a notion in which players can adapt portions of their full (normal-form) strategy over time, rather than reinforce or dispose of entire strategies in an all-or-nothing fashion.

While this means that we can no longer be agnostic about the way in which agents learn or evolve in these games, the inclusion of a weight parameter in each replicator equation allows us to more easily navigate between different versions of the model. Conveniently, we find that for certain classes of games, there is a natural setting of this parameter that yields back the original replicator dynamic over normal-form strategies. Even better, a second setting of the parameter results in backward induction reasoning — a process that is naturally associated with sequential games, but has previously been missing from evolutionary models. In this sense, the weight parameter can also be thought of as a degree of rationality we assign to agents — a higher weight corresponds to a higher rate of adaptation in rare subgames, i.e., a more rational agent.

There are of course many directions for future investigation of this model. For starters, the results here only focus on two classes of sequential games — the centipede game and the ultimatum game. While these games do in a sense cover the two types of relations between information sets in a game tree, it remains to be shown that the same equivalence result (Theorem 3) still holds in other games that combine dependent and independent information sets.

Additionally, there are other options that can be considered in place of the existing param-

eters given here. One naturally existing notion that our model does not address is that of analogical learning, in which an agent might use observations in one information set to adapt not only their local strategy there, but also their local strategies in other *similar* information sets as well. A thorny problem for this approach will be to define an appropriate notion of similarity.

In the context of simultaneous games, the replicator dynamic (and more generally, many evolutionary models) have a nice one-model-fits-all feature in that they can take on different interpretations, e.g., learning or evolution. Although potentially incomplete, our model shows that we can no longer do this with sequential games. One must determine how players learn or reason about information sets that are rarely (or never) reached in the play of the game. There are multiple reasonable answers to this question, depending on the intended interpretation. Which answer one chooses will have significant implications, including whether or not the players are expected to evolve toward the subgame perfect equilibrium.

7 Appendix

7.1 Definition of Extensive Form Games

An N -player extensive form game is a 4-tuple consisting of the following four components¹⁰:

1. **Game Tree:** The game tree is a finite tree starting at a single vertex called the *root*. The unique sequence of edges and vertices connecting the root to a vertex x is called the path to x . A vertex y is said to precede x if y is different from x and y is on the path to x . Likewise, y is said to follow x if y is different from x and x is on the path to y . A vertex z is an endpoint if no vertex follows z , and the set of all endpoints is denoted Z . A vertex that is not an endpoint is called a decision point, and the set of all decision points is denoted X . The edges connecting a decision point x to a vertex following x are called actions at x .
2. **Player Partition P :** The player partition P partitions the set of decision points X into $N + 1$ sets P_0, P_1, \dots, P_N , where P_0 is the set of decisions by nature and P_n is the set of decisions of player n . Because this thesis only considers perfect information games, P_0 is assumed to be empty.
3. **Information Partition U :** The information partition U is an N -tuple (U_1, \dots, U_N) , where each U_n partitions player n 's set of decisions P_n into information sets u such that each $u \in U_n$ satisfies the following two conditions:
 - (a) Every path intersects u at most once.
 - (b) All decision points in u have the same number of actions.

In perfect information games, all information sets consist of single decision points - this is equivalent to saying that there are no simultaneous moves by more than one

¹⁰The notation used here is adapted from Cressman (2003) and Kuhn (1953), although some parts have been simplified due to the restriction to generic perfect information games.

player at any point in the game. The set of actions available at an information set u is denoted C_u .

4. **Payoff Function π :** The payoff function is an N -tuple $\pi = (\pi_1, \dots, \pi_N)$, where each $\pi_n : Z \rightarrow \mathbb{R}$ assigns a real number payoff to player n at each endpoint z .

A pure strategy i of player n , denoted $e_{n,i}$, specifies a choice at each of the information sets of this player. In some of specific games below, pure strategies will sometimes be written in terms of these choices for convenience; for example, in the Mini Ultimatum game, the strategies of Player 2 are AA, AR, RA, RR .¹¹

A mixed strategy $p_n = (p_{n,1}, p_{n,2}, \dots)$ is a probability vector over the set of pure strategies, whose coordinates $p_{n,i}$ specify the probabilities of each pure strategy $e_{n,i}$ being used.

The normal form of a (2-player) extensive form game is a bimatrix game G consisting of the pure strategies of both players, and the payoff function π . Specifically, we have $G = (A, B)$, where A and B are the payoff matrices of each player with entries $a_{ij} = \pi_1(e_i, f_j)$ and $b_{ji} = \pi_2(e_i, f_j)$

7.2 Proofs

We begin with a lemma on writing the replicator dynamic equations for games in which one player has exactly two strategies. This lemma extends to the induced replicator dynamic as well, and is particularly useful there as all of the games we consider contain information sets of exactly two actions.

Lemma 5. Suppose that a population consists of just two strategies s_1 and s_2 (with proportions p_1 and p_2 , and fitnesses f_1 and f_2 respectively). Then, the replicator dynamic has the following equivalent form:

$$\dot{p}_1 = p_1(f_1 - \phi) = p_1(1 - p_1)(f_1 - f_2)$$

Proof: Note that the average fitness can be expressed in terms of the individual fitnesses: $\phi = \sum_i p_i f_i = p_1 f_1 + p_2 f_2 = p_1 f_1 + (1 - p_1) f_2$. Substituting this into the original equation above immediately gives the second form.

Elementary Perfect Information Games

An elementary perfect information game is defined to have the following extensive and normal forms:

Lemma 6. For all elementary perfect information games, the standard replicator dynamic and induced replicator dynamic using realized probabilities are equivalent.

Proof: Since Player 1's information set is first in the game, its realization probability is 1, and thus the two replicator dynamics are trivially equivalent. For Player 2, the standard

¹¹Mathematically, e_i is the i -th unit vector of length $|S_i|$

replicator dynamic gives

$$\dot{p}_l = p_l(1 - p_l)(1 - p_L)(b - c),$$

while the induced replicator dynamic with $K^u = 1 - p_L$ gives

$$\dot{p}_l = (1 - p_L)p_l(1 - p_l)(b - c).$$

Ultimatum Game (Theorem 2)

We begin this proof with some notation that will be useful later on:

Let $s \in S_1$ denote a strategy (offer) of Player 1, and let $t \in S_2$ denote a strategy of Player 2. As a bit of shorthand, let s also denote the payoff for Player 2 if she accepts offer s ; let r denote the rejection payoff (typically 0). Note that we are primarily interested in the equivalence of the new dynamic over Player 2 strategies in this game, so we can disregard Player 1's payoffs altogether.

Given an offer $s \in S_1$, let $A_s \subseteq S_2$ denote the set of Player 2 (normal-form) strategies that accept s , and let $R_s = S_2 \setminus A_s$ denote the strategies that reject s . Note that A_s also represents in the induced replicator dynamic the proportion of accepting offer s at that particular information set (i.e., we have a change of variables $p_{A_s} = \sum_{t \in A_s} p_t$). Similarly, given a strategy $t \in S_2$, let $O_t \subseteq S_1$ denote the offers that t accepts, and $N_t = S_1 \setminus O_t$ denote the offers that t rejects.

As an example, in the 3-offer Ultimatum game, Player 2 has 8 normal-form pure strategies of the form $\{A, R\}^3$. In the extensive form, Player 2 has three different information sets, so a behavioral strategy is a triple of probabilities $\{p_{A_1}, p_{A_2}, p_{A_3}\}$. The change of variables between these two forms is the following: for each local behavioral strategy A_s , we have $p_{A_s} = \sum_{t \in A_s} p_t$; for each pure strategy t , we have $p_t = \prod_{s \in O_t} p_{A_s} \prod_{s \in N_t} (1 - p_{A_s})$. An example in this game would be $p_{A_1} = p_{AAA} + p_{AAR} + p_{ARA} + p_{ARR}$, and $p_{AAR} = p_{A_1} p_{A_2} (1 - p_{A_3})$.

Fix some $s^* \in S_1$. We want to show that $\dot{p}_{A_{s^*}} = \sum_{t \in A_{s^*}} \dot{p}_t$ (where the LHS uses equations from the induced RD and the RHS from the standard RD).

Under the induced RD with realized probabilities, the LHS becomes

$$\dot{p}_{A_{s^*}} = p_{s^*} p_{A_{s^*}} (1 - p_{A_{s^*}}) (s^* - r).$$

On the RHS we have

$$\sum_{t \in A_{s^*}} \dot{p}_t = \sum_{t \in A_{s^*}} p_t (f_t - \phi) = \left(\sum_{t \in A_{s^*}} p_t f_t \right) - p_{A_{s^*}} \phi = \sum_{t \in A_{s^*}} p_t f_t - p_{A_{s^*}} \sum_{t \in S_2} p_t f_t.$$

The second summation can be broken into two summations over the sets of strategies that accept and reject s^* (A_{s^*} and R_{s^*}):

$$\begin{aligned} \sum_{t \in A_{s^*}} p_t f_t - p_{A_{s^*}} \sum_{t \in S_2} p_t f_t &= \sum_{t \in A_{s^*}} p_t f_t - p_{A_{s^*}} \sum_{t \in A_{s^*}} p_t f_t - p_{A_{s^*}} \sum_{t \in R_{s^*}} p_t f_t \\ &= (1 - p_{A_{s^*}}) \sum_{t \in A_{s^*}} p_t f_t - p_{A_{s^*}} \sum_{t \in R_{s^*}} p_t f_t. \end{aligned}$$

We now expand the fitness terms f_t within each summation. For each Player 2 strategy t , we have three different cases:

1. Player 1 offers s^* (in which case Player 2 receives either s^* or r depending on whether it's in A_{s^*} or R_{s^*}),
2. Player 1 offers some s other than s^* which is accepted by t (in which case Player 2 receives s)¹²
3. Player 1 offers some s other than s^* which is rejected by t (in which case Player 2 receives r).

$$(1 - p_{A_{s^*}}) \sum_{t \in A_{s^*}} p_t \left(\underline{p_{s^*} s^*} + \sum_{s \in O_t} p_s s + \sum_{s \in N_t} p_s r \right) - p_{A_{s^*}} \sum_{t \in R_{s^*}} p_t \left(\underline{p_{s^*} r} + \sum_{s \in O_t} p_s s + \sum_{s \in N_t} p_s r \right).$$

The first two terms of each summation (underlined above) can be combined to get

$$p_{s^*} p_{A_{s^*}} (1 - p_{A_{s^*}}) (s^* - r),$$

which is exactly the LHS in the original theorem. Thus, it suffices to show that the remaining terms,

$$(1 - p_{A_{s^*}}) \sum_{t \in A_{s^*}} p_t \left(\sum_{s \in O_t} p_s s + \sum_{s \in N_t} p_s r \right) - p_{A_{s^*}} \sum_{t \in R_{s^*}} p_t \left(\sum_{s \in O_t} p_s s + \sum_{s \in N_t} p_s r \right).$$

sum to zero. We do this via a change in the order of summation — the basic idea is that we look at each Player 1 offer $s \neq s^*$, and collect terms according to whether Player 2's strategy t accepts or rejects s and s^* (the initially fixed offer of the behavioral strategy we're comparing):

$$\begin{aligned} & \sum_{s \in S_1 \setminus s^*} p_s s \left((1 - p_{A_{s^*}}) \sum_{\substack{t \in \\ A_s \cap A_{s^*}}} p_t - p_{A_{s^*}} \sum_{\substack{t \in \\ A_s \cap R_{s^*}}} p_t \right) + p_s r \left((1 - p_{A_{s^*}}) \sum_{\substack{t \in \\ R_s \cap A_{s^*}}} p_t - p_{A_{s^*}} \sum_{\substack{t \in \\ R_s \cap R_{s^*}}} p_t \right). \\ &= \sum_{s \in S_1 \setminus s^*} p_s s (p_{R_{s^*}} p_{A_s} p_{A_{s^*}} - p_{A_{s^*}} p_{A_s} p_{R_{s^*}}) + p_s r (p_{R_{s^*}} p_{R_s} p_{A_{s^*}} - p_{A_{s^*}} p_{R_s} p_{R_{s^*}}) = 0. \end{aligned}$$

In the second line above, we simplified the inner summations using the change of variables from normal-form to behavioral strategies. Since the terms within each set of parentheses cancel out, we see that the remaining expression is indeed zero, which completes the proof.

¹²Note that here we've adjusted the notation slightly so that O_t and N_t denote the strategies *other than* s^* that are accepted and rejected by t .

Finally, let ϕ_u denote the average payoff of Player 1 in the subgame starting at node u . As an example, in the centipede game of length 4 (whose extensive and normal forms are given in Figure 4 below), we have $\phi_{u_2} = P_2x_2 + (1 - P_2)(Q_2y_2 + (1 - Q_2)x_3)$, which denotes the average payoff of Player 1 in the subgame starting at u_2 .

The average payoff ϕ_{u_i} is equal to the average fitness of the submatrix containing rows and columns greater than or equal to i , normalized by the total proportion of those strategies. A trivial but important instance of this is the fact that $\phi_{u_1} = \phi$, where the LHS is the newly defined average payoff in extensive form, and the RHS is the standard definition of average payoff in the replicator dynamic on the normal form. Additionally, we can write relations between the average payoffs of adjacent subgames, which we state in the following lemma:

Lemma 7. For all i ,

$$(1 - \sum_{j=1}^{i-1} p_j)(x_i - \phi_{u_i}) = (1 - \sum_{j=1}^i p_j)(x_i - \phi_{v_i}).$$

Proof: From the definition of ϕ_u and the extensive form of the Centipede game, we have

$$\phi_{u_i} = P_i x_i + (1 - P_i) \phi_{v_i} = \frac{p_i x_i + (1 - \sum_{j=1}^i p_j) \phi_{v_i}}{1 - \sum_{j=1}^{i-1} p_j},$$

which can be rearranged to give the equation above.

Theorem 8. The induced replicator dynamic using realized probabilities and standard replicator dynamic differ by a factor of $K_{u_i} = 1 - \sum_{j=1}^{i-1} p_j$ for the evolution of D_i . That is¹³,

$$\dot{P}_i = (1 - \sum_{j=1}^{i-1} p_j) \frac{dP_i}{dt}.$$

Proof: LHS:

$$\dot{P}_i = K P_i (1 - P_i) (x_i - \phi_{v_i}) = (1 - \sum_{j=1}^{i-1} p_j) (1 - \sum_{j=1}^{i-1} q_j) \frac{p_i (1 - \sum_{j=1}^i p_j)}{(1 - \sum_{j=1}^{i-1} p_j)^2} (x_i - \phi_{v_i})$$

¹³Note the difference between the terms \dot{P}_i , which denotes the time derivative given by the induced replicator dynamic, and $\frac{dP_i}{dt}$ which differentiates P_i as an expression of p_i and then substitutes in the respective time derivatives \dot{p}_i given by the standard replicator dynamic

an endpoint that has payoffs equal to those of the highest-paying action for the player acting at that information set. As the process can continue inductively up the full game tree, this shows that the only asymptotically stable point must be that of the subgame-perfect equilibrium.

8 References

- [1] A W Beggs. “On the Convergence of Reinforcement Learning”. In: *Journal of Economic Theory* 122 (2005), pp. 1–36.
- [2] Carl T. Bergstrom and Michael Lachmann. “Signalling among Relatives. I. Is Costly Signalling Too Costly?” In: *Philosophical Transactions of the royal Society of London B* 352 (1997), pp. 609–617.
- [3] Ross Cressman. *Evolutionary dynamics and extensive form games*. Economic learning and social evolution 5. Cambridge, Mass: MIT Press, 2003. ISBN: 978-0-262-03305-3.
- [4] John Gale, Kenneth G. Binmore, and Larry Samuelson. “Learning to Be Imperfect: The Ultimatum Game”. In: *Games and Economic Behavior* 8.1 (1995), pp. 56–90. ISSN: 10902473. DOI: 10.1016/S0899-8256(05)80017-X.
- [5] Nicola Gatti, Fabio Panozzo, and Marcello Restelli. *Efficient evolutionary dynamics with extensive-form games*. arXiv:1304.1456 [cs]. Apr. 2013. DOI: 10.48550/arXiv.1304.1456. URL: <http://arxiv.org/abs/1304.1456> (visited on 03/24/2025).
- [6] John C. Harsanyi and Reinhard Selten. *A General Theory of Equilibrium Selection in Games*. Cambridge: MIT Press, 1988.
- [7] Simon M. Huttegger. “Evolution and the Explanation of Meaning”. In: *Philosophy of Science* 74 (January 2007), pp. 1–27.
- [8] Simon M. Huttegger. “Generic Properties of Evolutionary Games and Adaptationism”. In: *Journal of Philosophy* 107.2 (2010), pp. 80–102.
- [9] Simon M. Huttegger. “ON THE RELATION BETWEEN GAMES IN EXTENSIVE FORM AND GAMES IN STRATEGIC FORM”. In: *Reduction - Abstraction - Analysis*. Ed. by Alexander Hieke and Hannes Leitgeb. De Gruyter, Dec. 31, 2009, pp. 377–388. ISBN: 978-3-11-032887-5. DOI: 10.1515/9783110328875.377. URL: <https://www.degruyter.com/document/doi/10.1515/9783110328875.377/html> (visited on 04/29/2025).
- [10] H. W. Kuhn. “Extensive Games and the Problem of Information”. In: *Contributions to the Theory of Games (AM-28), Volume II*. Ed. by H. W. Kuhn and A. W. Tucker. Princeton University Press, 1953, pp. 193–216. ISBN: 978-0-691-07935-6. URL: <http://www.jstor.org/stable/j.ctt1b9x1zv.17> (visited on 05/28/2025).
- [11] Karl H. Schlag. “Why Imitate, and If So, How?” In: *Journal of Economic Theory* 78.1 (Jan. 1998), pp. 130–156. ISSN: 00220531. DOI: 10.1006/jeth.1997.2347. URL: <http://linkinghub.elsevier.com/retrieve/pii/S0022053197923474>.

- [12] R. Selten. “Reexamination of the perfectness concept for equilibrium points in extensive games”. In: *International Journal of Game Theory* 4.1 (Mar. 1, 1975), pp. 25–55. ISSN: 1432-1270. DOI: 10.1007/BF01766400. URL: <https://doi.org/10.1007/BF01766400> (visited on 05/28/2025).
- [13] Brian Skyrms. “Darwin Meets the Logic of Decision: Correlation in Evolutionary Game Theory”. In: *Philosophy of Science* 61.4 (1994), pp. 503–528. ISSN: 00318248. JSTOR: 188332. URL: <http://www.jstor.org/stable/188332>.
- [14] Brian Skyrms. *The Dynamics of Rational Deliberation*. Cambridge: Harvard University Press, 1990.
- [15] P Taylor and L Jonker. “Evolutionarily Stable Strategies and Game Dynamics”. In: *Mathematical Biosciences* 40 (1978), pp. 145–156.
- [16] Bernhard Von Stengel. “Efficient Computation of Behavior Strategies”. en. In: *Games and Economic Behavior* 14.2 (June 1996), pp. 220–246. ISSN: 08998256. DOI: 10.1006/game.1996.0050. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0899825696900500> (visited on 03/24/2025).
- [17] John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton: Princeton University Press, 1953.
- [18] Michael Weisberg. *Simulation and Similarity*. Oxford: Oxford University Press, 2013.
- [19] Kevin J.S. Zollman. “Explaining Fairness in Complex Environments”. In: *Politics, Philosophy, and Economics* 7.1 (2008), pp. 81–97.