Prioritised Learning in Snowdrift-Type Games

Maria Kleshnina [1],* , **Sabrina S. Streipert [2]**, **Jerzy A. Filar [3]** and **Krishnendu Chatterjee [1]**

[1] Institute of Science and Technology Austria (IST Austria), Am Campus 1, 3400 Klosterneuburg, Austria; krishnendu.chatterjee@ist.ac.at

[2] Department of Mathematics and Statistics, McMaster University, 1280 Main St W, Hamilton, ON L8S 4L8, Canada; streipes@mcmaster.ca

[3] Centre for Applications in Natural Resource Mathematics, School of Mathematics and Physics, University of Queensland, St Lucia, QLD 4072, Australia; j.filar@uq.edu.au

* Correspondence: maria.kleshnina@ist.ac.at

check for updates

Abstract: Cooperation is a ubiquitous and beneficial behavioural trait despite being prone to exploitation by free-riders. Hence, cooperative populations are prone to invasions by selfish individuals. However, a population consisting of only free-riders typically does not survive. Thus, cooperators and free-riders often coexist in some proportion. An evolutionary version of a Snowdrift Game proved its efficiency in analysing this phenomenon. However, what if the system has already reached its stable state but was perturbed due to a change in environmental conditions? Then, individuals may have to re-learn their effective strategies. To address this, we consider behavioural mistakes in strategic choice execution, which we refer to as incompetence. Parametrising the propensity to make such mistakes allows for a mathematical description of learning. We compare strategies based on their relative strategic advantage relying on both fitness and learning factors. When strategies are learned at distinct rates, allowing learning according to a prescribed order is optimal. Interestingly, the strategy with the lowest strategic advantage should be learnt first if we are to optimise fitness over the learning path. Then, the differences between strategies are balanced out in order to minimise the effect of behavioural uncertainty.

Keywords: incompetence; evolutionary games; learning; cooperation; snowdrift game

1. Introduction

In many situations, ability to learn becomes important for survival: how fast and effectively individuals explore their surroundings and learn how to react to them determines their survival chances. Some environments and situations might require specific skills or rapid adaptations whereas others might just require general awareness. Learning and adaptation are remarkable features of life in natural environments that are prone to frequent changes. The process of learning can be considered as a continuous multidimensional process where several skills can be learned in parallel. However, reducing the time step to a very small interval, we focus on learning only one skill at a time. In this study, we introduce a model of multidimensional stepwise learning by assuming that the learning process can be subdivided into smaller time steps during which we only learn one skill.

Evolutionary games were proposed to answer questions of the effects of natural selection on populations. Here, a game happens at a higher level: the game is not played directly by individuals but rather by strategies or skills competing for survival [1,2]. In such settings, learning the best and more effective strategy first becomes critical for individuals, especially under resource limitation [3]. Learning

in the evolutionary settings often assumes that learning evolves in parallel with the evolution itself [4]. In this paper, we consider the case when the system is perturbed after reaching its stable equilibrium. Finding the most efficient learning path to return to the previous equilibrium is a challenge. In this paper, we suggest to look at this problem from the evolutionary perspective where an optimal learning path is dictated by the chance of survival. This is achieved through maximisation of fitness depending on the learning path an individual takes.

Behavioural stochasticity is a popular object of study for game-theorists. First, the concept of "trembling hands" [5] was suggested as an approach to model players' mistakes during the strategies' execution with some small probability. In addition, mutations can be interpreted as reactions to environmental changes (see for instance [6–8] for studies on mutations). In artificial intelligence and economics this idea is of a particular interest since individuals can be imperfect and incapable of executing their strategies (examples of such studies are [9–11]). However, we do not restrict our analysis to mistakes that are random noise, but, instead, allow for biased mistakes and assume that skills can be learnt at different rates. For this, we utilise the notion of incompetence that means that individuals make mistakes with certain probabilities, which was first introduced in [12,13] and extended in [14]. Hence, learning is defined as a process of improving competence and reducing the probabilities of mistakes. Here, we propose the notion of prioritised learning when the priority in the order is determined by the skills' relative advantages. That is, we are interested in the mechanism of discovering an optimal order according to which skills should be learned. This mechanism relies on the interplay between relative fitness and learning advantages and the notion of prioritised learning which aims to balance out these differences between the strategies.

We focus on the question in which order species should learn their skills to have an evolutionary advantage. The process of re-learning the effective strategy can be considered as a continuous multidimensional problem. However, reducing the time step to a very small interval, we focus on only one skill. We consider the case of the coexistence of two strategies: to cooperate and to defect. Mathematically, this can be described by the Snowdrift (also known as Chicken or Hawk-Dove) game where two strategies interact and both appear in the behavioural traits of individuals [15].

The question of how cooperation evolves in communities is a rich field of study for mathematicians and biologists. Enforcement, punishment and reciprocity are the mechanisms that can sustain stable cooperation [16–19]. The Prisoners' Dilemma is the well-known example of the problem of cooperation [20,21]. It is a strict game where sustaining cooperation becomes a challenge. However, it can be relaxed if we assume that benefit gained from cooperation can be received by both players and the costs shared (see [22] for the review). The resulting game has a form of the Snowdrift game, where a stable mixed equilibrium exists [23]. We point out that the focus of our study is not the evolution of cooperative mechanisms as such. Learning to cooperate is a fruitful field proposing many interesting results [24–26]. However, here we are more interested in understanding of how the system should optimally evolve back to an equilibrium state once it was disturbed. Specifically, we shall focus on a Snowdrift type game in order to determine which of the strategies (cooperation or defection) should be learnt first.

When speaking of the optimality of a learning path, we expect that the benefit from this strategy has an impact on the overall population's fitness [27]. We note that learning and replicator timescales are decoupled. In fact, we require that the learning timescale is much slower than reproduction. This can be seen in view of a behavioural adaptation that might take longer time to happen. In classic settings of the replicator dynamics, the dynamics' time scale is equal to the reproduction time scale. That is, every time step is the length of one generation. However, when modelling behaviour of individuals, this reproduction time scale represents an interaction time scale. Therefore, the behavioural change might take several interactions to be achieved. Hence, naturally, we assume that individuals are learning slower than they

interact. For example, in the incompetent version of the Snowdrift game, individuals go through multiple interactions which may change their behaviour from cooperating to defecting.

We show that measuring the fitness over the learning path depends on the order of learning along the path and the extent to which strategies are learnt. Our results demonstrate that it might be preferable to learn a skill prone to higher probabilities of mistakes first and leverage its learning advantage. This suggests that in the environmental or other changes those strategies that are most disruptive must be adapted to more quickly if they are to survive at all. If two skills are equivalent in their relative strategic advantages, we show that both skills are also equivalent with respect to the order of learning. Counter-intuitively, we show that these relative advantages can still be identical even if the relative fitness advantages of the skills are significantly different suggesting that the evolution is trying to balance out mistakes. We conjecture that not only the fitness of the skill has to be taken into account but also the degree of incompetence.

In the following section we set up the model and define the notions of relative fitness, learning and strategic advantages. Then, in Section 3 we define a fitness-over-learning objective function measuring how fitness of the population improves over the learning path taken. After that, we proceed to two cases: (a) the case when skills are identical in their strategic advantage and no prioritised learning is needed in Section 4, and (b) the case when skills are distinct in their advantages in Section 5. These two cases demonstrate the notion of hierarchy in the learning order and how mistakes affect the evolution.

2. Learning Setup

Typically, a population of species acquires a set of skills that they are required to learn either while being young or while adapting to new environmental conditions. Let us suppose that this set consists of only two essential skills both of which are needed for survival (or for stable coexistence in the population). Hence, we consider two specific strategies that need to be learnt: cooperation and defection. In the evolutionary context, defection can be interpreted as aggression to capture (rather than share) a resource such as food or territory. We utilise the notion of replicator dynamics in order to describe the evolution of interactions among individuals in the population [28,29]. Let $p := p(t) \in [0,1]$ be the frequency of cooperators at (evolutionary) time t. Then, the dynamics can be expressed as

$$\dot{p} = (f_C - \phi)p, \tag{1}$$

where f_C is fitness of cooperators and ϕ is the mean fitness function defined as $\phi := f_C p + f_D(1-p)$, since, $(1-p)$ is the fraction of defectors. For the purpose of this paper, we use a linear form of the fitness functions, as in [14], which simplifies to

$$\dot{p} = (f_C - \phi)p = (f_C - f_D)p(1-p). \tag{2}$$

We shall note that if both cooperation and defection strategies would be required to some extent, then both strategies might be required to coexist at equilibrium. Such an equilibrium usually characterises the Snowdrift game, which is also referred to as a anti-coordination game. Since both strategies are the best response to the opposite strategy, at equilibrium, both strategies coexist, hence, securing some stable level of cooperation. We shall construct a reward matrix R for such a game as

$$
\begin{array}{c}
\begin{array}{cc} \text{Cooperate} & \text{Defect} \end{array} \\
\begin{array}{c} \text{Cooperate} \\ \\ \text{Defect} \end{array}
\begin{pmatrix} B - C/2 & B - C \\ \\ B & 0 \end{pmatrix}
\end{array}
$$

where B is the benefit and C is the cost of cooperation. In order to simplify our analysis, we apply the linear transformation from [30] and subtract the diagonal elements from the corresponding columns, as it does not affect dynamics' behaviour. Then, we can consider a reduced form of the matrix given by

$$
\begin{array}{cc}
 & \text{Cooperate} \quad \text{Defect} \\
\begin{array}{c} \text{Cooperate} \\[1em] \text{Defect} \end{array} &
\begin{pmatrix} 0 & a \\ b & 0 \end{pmatrix},
\end{array}
\tag{3}
$$

where $a = B - C$ and $b = C/2$. As we want to set our game to be a Snowdrift game, we assume that $B > C > 0$, because then both strategies will stably coexist at the equilibrium $\hat{\mathbf{p}} = (\hat{a}, \hat{b})$ and their frequencies will be given by

$$
\hat{a} := \frac{a}{a+b} = \frac{2B - 2C}{2B + C} \quad \text{and} \quad \hat{b} := \frac{b}{a+b} = \frac{C}{2B + C}.
\tag{4}
$$

Then, in the context of learning a set of skills, both skills will be required to be learned. The question we address is: which one should be learnt first?

In order to model learning these skills in the game, we utilise the notion of incompetence. This concept was first introduced for classic games in 2012 [12] and extended to evolutionary games in 2018 [14]. Here, individuals choose a strategy to play but have non-zero probabilities of executing a different strategy. Such mistakes are a manifestation of the incompetence of players. Mathematically, it is described by the matrix of incompetence, Q, that evolves as individuals are learning. This matrix consists of conditional probabilities q_{ij} determining the probability of executing strategy j given that strategy i was chosen. The matrix has the form

$$
Q = \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix}.
$$

A schematic representation of the interaction under incompetence can be found in Figure 1A. As a measure of learning we use parameters x and y for each strategy. Thus, each strategy can be learned at a different pace, which sets this work apart from the existing literature (see [12,31]). Then, the matrix of incompetent parameters has the form

$$
X = \begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix},
$$

where $x, y \in [0, 1]$. We assume that strategies can be learnt and, hence, the incompetence can decrease from some initial level of propensity to make mistakes. Thus, the learning process is described by the equation

$$
Q(X) = (I - X)S + X,
\tag{5}
$$

where S represents the starting level of incompetence. If $x = y = 0$, then $X = \mathbf{0}$ and

$$
Q(\mathbf{0}) = S = \begin{pmatrix} \alpha & 1 - \alpha \\ 1 - \beta & \beta \end{pmatrix}.
\tag{6}
$$

Full competence corresponds to the case $x = y = 1$, where $Q = I$, the identity matrix. Each strategy has its own measure of incompetence level and its own time needed for it to be mastered. Then, the new incompetent game reward matrix is defined as

$$R(X) = R(x, y) = Q(X) \times R \times Q(X)^T.$$

Given the new reward matrix, we require some basic assumptions on parameters in (3)–(5) to be satisfied in order for the game to avoid bifurcations in the replicator dynamics. Specifically:

1.　$a, b > 0$: this ensures that it is beneficial to learn both strategies.
2.　$\alpha > \hat{a}$, $\beta > \hat{b}$: this is necessary for the new incompetent game to have a stable mixed equilibrium point for any level of incompetence, $\forall x, y \in [0, 1]$ [14].
3.　$\alpha + \beta > 1$: this condition also guarantees existence of the stable mixed equilibrium [14].

If parameters of the game do not meet these conditions, then there will be some values of the incompetence parameters for which the system (1) undergoes bifurcation. This would lead to situations where one of the skills is not beneficial to be learnt due to the fact that it is dominated. Hence, learning the beneficial skill would be an obvious answer to the question of which skill to learn first. However, we are interested in the case when the optimal learning path involves both skills.
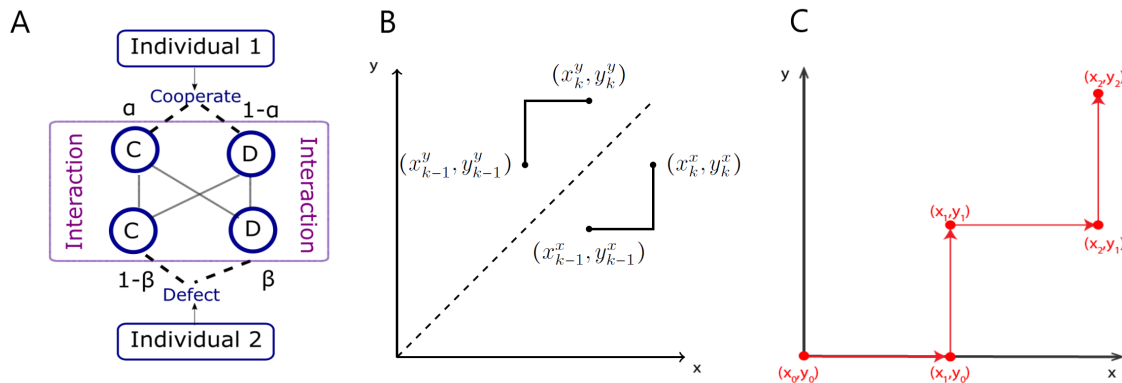

Figure 1. (**A**) A schematic representation of the effect of mistakes: individual 1 chooses to cooperate and individual 2 chooses to defect. However, under incompetence, individual 1 will cooperate only with probability α and individual 2 will defect only with probability β. Hence, the outcome of the interaction now depends on the probability distribution of mistakes. (**B**) A schematic representation of prioritised learning: we aim to define first an optimal direction in which the learning should start (x and y direction represents strategies 1 and 2, respectively) and an n-step optimal sequence of (x_k^x, y_k^x) if x direction is optimal (or (x_k^y, y_k^y) if y direction is optimal). (**C**) An example of a learning path in the x direction. We start at (x_0, y_0), then taking the first half-step in the x direction to (x_1, y_0) and a second half-step in the y direction to (x_1, y_1). By design, we have to start at $(0, 0)$ and finish at $(1, 1)$.

3. Maximising Fitness over Learning

The optimality of a learning path can be determined in many ways. In the replicator dynamics with symmetric payoff matrices, interactions follow the evolutionary path that maximises the overall population's fitness [32]. We focus on the fitness function at the equilibrium state, which implies that the population attains a steady-state faster than incompetence parameters change. Technically speaking, we assume either very long timescales of learning or else very fast convergence to steady state, or both.

Hence, it is sufficient to consider the mean fitness function [29] of the Snowdrift game which has the following form

$$\phi = \hat{\mathbf{p}} R \hat{\mathbf{p}}^T = \frac{ab}{a+b}.$$

Then, in our new incompetent game the mean fitness $\phi(X) = \hat{\mathbf{p}} R(X) \hat{\mathbf{p}}^T$ can be shown to be

$$\phi(X) = (a+b)\Big\{(\alpha - \hat{a})(\beta - \hat{b}) + y(1-\beta)\alpha - \hat{a} + x(1-\alpha)(\beta - \hat{b}) + xy(1-\alpha)(1-\beta)\Big\}.$$

We shall note that in fact any nontrivial learning path will improve the mean fitness at the equilibrium. This follows from the fact that for any vector $\mathbf{x} = (x, y)$ with entries in $(0, 1)$ the following holds

$$\mathbf{x} H_\phi \mathbf{x}^T = 2(a+b)xy(1-\alpha)(1-\beta) \geq 0,$$

where H_ϕ is a Hessian of $\phi(X)$. Since we would like to find an optimal learning path that maximises the fitness, it is convenient to consider a new re-scaled fitness function. For that, we choose

$$\tilde{\phi}(X) := \frac{\phi(X)}{(a+b)(1-\alpha)(1-\beta)} = \tilde{a}\tilde{b} + \tilde{b}x + \tilde{a}y + xy, \tag{7}$$

where new parameters $\tilde{a}$ and $\tilde{b}$ are

$$\tilde{a} := \frac{\alpha - \hat{a}}{1 - \alpha}, \quad \tilde{b} := \frac{\beta - \hat{b}}{1 - \beta}. \tag{8}$$

An important note for further analysis is the understanding of fitness and learning advantages. Given that the relative fitness of each strategy is positive, that is, $a, b > 0$, we say that the strategy with a higher relative fitness obtains a fitness advantage. In addition, we define a learning advantage of a strategy in a very special way. This advantage arises when the strategy is accompanied by a higher probability of making a mistake in its execution. In such a case reducing the corresponding level of incompetence offers a greater opportunity for improvement, thereby constituting a learning advantage.

Next, note that the parameters $\tilde{a}$ and $\tilde{b}$ introduced in (8) are closely connected with the above definitions of fitness and learning advantages. Indeed, their difference allows us to capture the relative tradeoffs between these two types of advantages. Hence, we shall say that the new parameter $\delta := \tilde{a} - \tilde{b}$ measures the relative strategic advantage of cooperators against defectors. That is, if $\delta > 0$, cooperators have an advantage. We illustrate these concepts in Table 1. Next, we shall demonstrate how this notion affects the optimal learning path with respect to the population's mean fitness.

Table 1. Definitions of advantages of cooperators over defectors. For the definition of advantages of defectors over cooperators, the inequality signs in parameters comparison should be reversed.

Advantage	Parameters Comparison	Discussion
Fitness advantage	$a > b$	Fitness advantage implies that cooperator have higher fitness and, hence, are more abundant.
Learning advantage	$\alpha < \beta$	Learning advantage implies that cooperators are more flexible in their strategy execution and can act as defectors.
Strategic advantage	$\delta > 0$ or $\tilde{a} > \tilde{b}$	Strategic advantage combines both fitness and learning advantages implying that if cooperators are disadvantaged in fitness (or learning), then advantage in learning (or fitness) can compensate.

However, we need to take into account possible limitations of the learning pace. We assume that only one skill can be learnt to some degree and two skills cannot be learnt simultaneously. Thus, learning is achieved in a discrete manner (see Figure 1B). The question is what is the optimal learning order and what are the switching points. This will be determined by defining the learning path as a curve $\mathcal{C}$ on the learning space $\mathcal{L}$ such that it starts at $(0,0)$ and ends at $(1,1)$.

Definition 1. *Define the* learning space *of an incompetence game* $\mathcal{L}$ *as the domain of the incompetence matrices* $Q(X)$ *from* (5) *given by the set of all* 2×2 *stochastic matrices.*

Definition 2. *Define a* learning path *for an incompetence game as a curve* $\mathcal{C} : I^2 \to \mathcal{L}$ *on the learning space* $\mathcal{L}$ *such that* $\mathcal{C}(\mathbf{0}) = \mathbf{0} = (0,0)^T$ *and* $\mathcal{C}(\mathbf{1}) = \mathbf{1} = (1,1)^T$.

A learning path $\mathcal{C}$ can be a smooth curve or a stepwise path, depending on whether learning is continuous or discrete. We shall consider only stepwise learning paths. This is a natural restriction because: at a small time-scale only one skill can be learned at any given time. However, we will also study the case when the number $n \to \infty$, which approaches a smooth learning curve.

Definition 3. *A stepwise learning path of order* n, $\mathcal{C}_n = \{(x_k, y_k) : 1 \le k \le n\}$, *is a stepwise curve in the learning space* $\mathcal{L}$, *connecting the n points* $(x_k, y_k)_{k=1}^n$, *that satisfies*

(a) $x_{k+1} > x_k, \quad 1 \le k \le n-1,$
(b) $y_{k+1} > y_k, \quad 1 \le k \le n-1,$
(c) $x_1 = y_1 = 0,$
(d) $x_n = y_n = 1.$

Conditions (a) and (b) imply $\Delta x_k, \Delta y_k > 0$ for $1 \le k \le n-1$. The path segment from (x_k, y_k) to (x_{k+1}, y_{k+1}) could consist of the following sequence $(x_k, y_k), (x_{k+1}, y_k), (x_{k+1}, y_{k+1})$, in which case we say that the x direction was taken. The alternative path segment is $(x_k, y_k), (x_k, y_{k+1}), (x_{k+1}, y_{k+1})$, indicating that the y direction was chosen first.

Definition 4. *The set of all (alternating) stepwise curves of order* n *that satisfy conditions (a)–(d) is called the* learning set of order n *and is denoted by* $\mathcal{L}_n$.

Here, we focus on alternating stepwise paths where the first direction determines the remaining path. Consequently, an n-step learning path is described by the points $\{(x_k, y_k)\}_{k=1}^n$ satisfying (a)–(d) above, resulting in two possible learning paths corresponding to the direction of the first step. Let $\Phi_x(\mathcal{C}_n)$ be the fitness-over-learning in the x direction for the learning path $\mathcal{C}_n \in \mathcal{L}_n$ given by

$$\Phi_x(\mathcal{C}_n) = \sum_{k=1}^{n-1} \left\{ \int_{x_k}^{x_{k+1}} \phi(s, y_k)\, ds + \int_{y_k}^{y_{k+1}} \phi(x_{k+1}, s)\, ds \right\}. \tag{9}$$

Let $\Phi_y(\mathcal{C}_n)$ be the fitness-over-learning in the y direction for the learning path $\mathcal{C}_n \in \mathcal{L}_n$ defined as

$$\Phi_y(\mathcal{C}_n) = \sum_{k=1}^{n-1} \left\{ \int_{y_k}^{y_{k+1}} \phi(x_k, s)\, ds + \int_{x_k}^{x_{k+1}} \phi(s, y_{k+1})\, ds \right\}. \tag{10}$$

That is, for a given $n \in \mathbb{N}$, we have two objective functions: Φ_x and Φ_y. Finding their maxima separately yields an optimal learning path with the optimal direction. In what follows, we define the

optimal learning paths in the x and y directions and the overall optimal learning path that maximises the population's measure of fitness over learning.

Definition 5. *The optimal learning path $\mathcal{C}_n^* \in \mathcal{L}_n$ with respect to the population's mean fitness function $\phi(X) : \mathcal{L}_n \to \mathbb{R}$ is such that it satisfies the equation*

$$\mathcal{C}_n^* = \begin{cases} \mathcal{C}_n^{*x}, & if \quad \Phi_x(\mathcal{C}_n^{*x}) > \Phi_y(\mathcal{C}_n^{*y}) \\ \mathcal{C}_n^{*y}, & if \quad \Phi_x(\mathcal{C}_n^{*x}) < \Phi_y(\mathcal{C}_n^{*y}) \end{cases}$$

where $\mathcal{C}_n^{*x}, \mathcal{C}_n^{*y}$ are the optimal paths in the x and y direction, respectively. When $\Phi_x^* = \Phi_y^*$, both $\mathcal{C}_n^{*x}$ and $\mathcal{C}_n^{*y}$ are optimal paths, where $\Phi_j^* = \Phi_j^*(\mathcal{C}_n^{*j}), j \in \{x, y\}$.

The superscript (respectively, subscript) indicates the direction of the path. That is, if the x direction is optimal, that is $\Phi_x^* > \Phi_y^*$, then $\mathcal{C}_n^* = \mathcal{C}_n^{*x}$. Otherwise, if $\Phi_x^* < \Phi_y^*$, then $\mathcal{C}_n^* = \mathcal{C}_n^{*y}$. When $\Phi_x^* = \Phi_y^*$, both directions x and y are optimal and it does not matter which one we take. Hence, we can define prioritised learning in these settings.

Definition 6. *We say that there exists prioritised learning for Φ among stepwise learning paths of order n, if there exists $\mathcal{C}_n^*$ such that one of the directions is preferable over the other. That is, $\Phi_x(\mathcal{C}_n)^* \neq \Phi_y(\mathcal{C}_n)^*$.*

Given the structure of the fitness Functions (7) and (9)–(10) we can explicitly derive Φ_x and Φ_y for the learning paths in the x and y directions. However, we will show that the optimal direction of learning can be determined simply by the sign of the relative strategic advantage, δ.

4. No Strategic Advantages

Throughout this section, we assume that no strategy has a relative strategic advantage ($\delta = 0$). We shall first note that in this case both objective functions, Φ_x and Φ_y, exhibit a symmetry relation.

Theorem 1. *If $\delta = 0$, then there is no difference in the direction of optimal learning, that is, $\Phi_x^* = \Phi_y^*$.*

Hence, if there is no relative strategic advantage in the game ($\delta = 0$), the order of learning does not affect the fitness of the population. It is therefore sufficient to calculate only one path that maximises the fitness-over-learning of the population. In the following proposition we show that this learning path has a remarkably simple form.

Proposition 1. *If $\delta = 0$, then the unique optimal stepwise learning path of order n in the x direction, $\mathcal{C}_n^{*x} = \{(x_k^{*,x}, y_k^{*,x}) : 1 \leq k \leq n\} \in \mathcal{L}_n$, is given by*

$$x_k^{*,x} = \begin{cases} 0, & k = 1 \\ \frac{2k-3}{2n-3}, & 2 \leq k \leq n \end{cases}, \qquad y_k^{*,x} = \begin{cases} \frac{2k-2}{2n-3}, & 1 \leq k \leq n-1 \\ 1, & k = n \end{cases}. \tag{11}$$

See Mathematical Appendix A for the proofs of Theorem 1 and Proposition 1. Interestingly, the optimal solution for the x direction yields $x_k^{*,x}, y_k^{*,x}$ such that $y_k^{*,x} \geq x_k^{*,x}, \forall k$. That is, each step in the y direction, y_k is greater than the step in the x direction, x_k.

To analyse how increasing the number of steps is changing the objective function, we consider the rate of change of the fitness over learning function at the optimal solution, that is

$$\Delta \Phi_x^* = \Phi_x^*(n+1) - \Phi_x^*(n). \tag{12}$$

After substituting the optimal solution (11) into (12) we obtain

$$\Delta\Phi_x^* = \frac{4}{3(2n-3)} > 0, \forall n > 2.$$

Due to symmetry, when $\delta = 0$, it follows that $\Delta\Phi_x^* = \Delta\Phi_y^*$.

Therefore, the smaller the learning steps, the greater the benefit. However, the marginal increases tend to 0 as n becomes large. Arguably, this illustrates the "law of diminishing returns" of stepwise learning.

Let us consider two following games, which will be called Examples 1 and 2:

$$R_1 = \begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix}, \quad S_1 = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix} \quad \text{and} \quad R_2 = \begin{pmatrix} 0 & 9 \\ 1 & 0 \end{pmatrix}, \quad S_2 = \begin{pmatrix} 0.9 & 0.1 \\ 0.9 & 0.1 \end{pmatrix}.$$

In terms of fitnesses, these examples are different: the fixed points of the replicator equations are $(0.5, 0.5)$ and $(0.9, 0.1)$, respectively. Hence, fitness of strategy 1 is higher in the second game. However, setting $\alpha_1 = \beta_1 = 0.5$ and $\alpha_2 = 0.9, \beta_2 = 0.1$, we have relative strategic advantages in examples 1 and 2 both equal to 0 ($\delta_1 = \delta_2 = 0$). Then, S_1 and S_2 equalise the strategies. The high probability of mistakes for strategy 2 in R_2 signals that it is more disrupted by incompetence. This makes the optimal learning paths for these two games identical (see Figure 2A).

Next we shall consider the case when $\delta \neq 0$. In this case, one of the strategies has a relative strategic advantage, depending on the sign of δ. We show that the order of the learning path is now important and influences the value of the fitness-over-learning function, which we call prioritised learning.
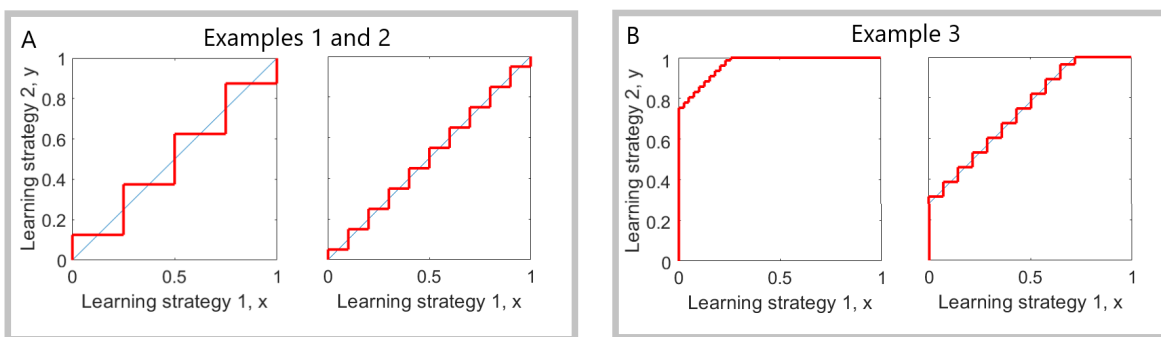

Figure 2. Optimal learning paths for examples in the manuscript. (**A**) Examples 1 and 2: optimal learning paths for R_1 and R_2 follow the line $x = y$ (blue line). We plot the solution (red stepwise line) for: (1) $n = 5$, (2) $n = 10$. (**B**) Example 3: an optimal learning path for R_3 where $n = 12$ and: (1) $a = 5$, (2) $a = 7$.

5. Prioritised Learning

First, we recall the notion of prioritised learning used henceforth. By Definition 6, there exists prioritised learning for Φ among stepwise learning paths of order n, if there exists $\mathcal{C}_n^*$ such that one of the directions is preferable over the other, along that path. That is, $\Phi_x(\mathcal{C}_n)^* \neq \Phi_y(\mathcal{C}_n)^*$. We shall next characterise an optimal solution in the x direction.

Proposition 2. *Let $n \in \mathbb{N}, n \geq 2$. If n is such that $\delta < \frac{1}{2(n-2)}$, then the x direction is preferred and the optimal learning path of order n is given by*

$$x_k^{*,x} = \begin{cases} 0, & k = 1 \\ -\delta + (1+\delta)\frac{2k-3}{2n-3}, & 2 \leq k \leq n \end{cases}, \quad y_k^{*,x} = \begin{cases} (1+\delta)\frac{2k-2}{2n-3}, & 1 \leq k \leq n-1 \\ 1, & k = n \end{cases}. \tag{13}$$

We refer the reader to Mathematical Appendix A for more details. Next, we provide conditions for which the y direction defines the optimal learning path.

Proposition 3. *Let $n \in \mathbb{N}, n \geq 2$. If n is such that $\delta > -\frac{1}{2(n-2)}$, then the y direction is preferred and the optimal learning path of order n is given by*

$$y_k^{*,y} = \begin{cases} 0, & k = 1 \\ \delta + (1-\delta)\frac{2k-3}{2n-3}, & 2 \leq k \leq n \end{cases}, \quad x_k^{*,y} = \begin{cases} (1-\delta)\frac{2k-2}{2n-3}, & 1 \leq k \leq n-1 \\ 1, & k = n \end{cases}. \tag{14}$$

Hence, depending on the value of δ, the optimal learning path has different directions. The threshold for δ is equal to $\left| \frac{1}{2(n-2)} \right|$, which for sufficiently large number of steps is nearly 0. However, in the proof of Theorem 2 (see Mathematical Appendix A), we show that it is the sign of δ that determines the direction of learning.

Theorem 2. *The direction of the optimal learning path is determined by the sign of δ: for $\delta > 0$ the y direction is optimal and for $\delta < 0$ the x direction is optimal.*

For $\delta = 0$ the optimal learning path represents $n - 2$ equally distributed steps and the direction of the first step does not affect the fitness-over-learning. Note that if $\delta \neq 0$, the first and last steps for each direction are

$$y_2^{*,y} = \delta + \frac{1-\delta}{2n-3}, \quad x_{n-1}^{*,y} = 1 - y_2^{*,y};$$

$$x_2^{*,x} = -\delta + \frac{1+\delta}{2n-3}, \quad y_{n-1}^{*,x} = 1 - x_2^{*,x}.$$

Thus, the first step of the learning path aims to adjust fitness and learning advantages between the two strategies. However, in the interior of the learning space, the path still takes $n - 2$ equally distributed steps. In Figure 3 we displayed the objective function Φ for different positive values of δ and different number of steps n. The images below the colormap enhance the changes in Φ^* as the number of steps n varies. The difference in the values of Φ^* with respect to n is marginal, hence, below we zoom into several values of δ. For $\delta < 1$, Φ_y^* increases in n, suggesting that taking more steps is beneficial. However, Φ_y^* decreases for $\delta > 1$, suggesting one-step learning of the skill. We show this in the next result that follows immediately from Propositions 2, 3, and Theorem 2.
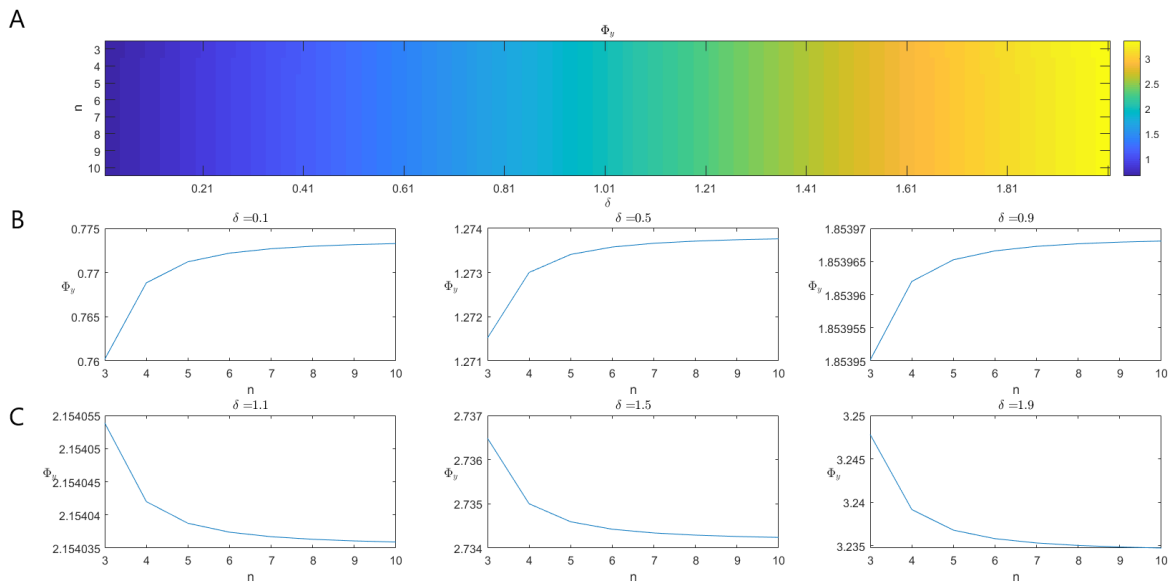
Figure 3. The fitness-over-learning Φ_y^* calculated for the optimal solution from (14) for different values of $\delta > 0$ and different step size n. (**A**) The colormap represents dependence of Φ_y^* on the number of steps and δ. (**B**) We plot the exact values of Φ_y^* for three values of $\delta \in (0, 1)$. As can be seen, the increasing number of steps improves the values of the fitness-over-learning function. (**C**) However, applying the same solution for higher values of the relative strategic advantage ($\delta > 1$) reduces the value of the objective function, see Corollary 1.

Corollary 1. *For* $|\delta| > 1$, *we obtain two cases:*

(i)   *If* $\delta \geq 1$, *then the optimal learning curve is a one-step function in the* y *direction.*

(ii) *If* $\delta \leq -1$, *then the optimal learning curve is a one-step function in the* x *direction.*

If $|\delta| > 1$, then the optimal solution from (13) and (14) is only feasible for $n = 2$. However, if δ is positive and less than 1, the greatest fitness-over-learning is achieved for a smooth learning path along the line $y = \delta + x$. This can be seen as a consequence of allowing n to approach infinity.

Corollary 2. *For* $0 < \delta < 1$, *the optimal step-wise learning path* $\{(x_k^{*,y}, y_k^{*,y})\}_{i=1}^n$ *in the* y *direction as* $n \to \infty$, *follows the relation*

$$
y^{*,y} = \begin{cases} \delta, & x^{*,y} = 0 \\ \delta + x^{*,y}, & x^{*,y} \in (0, 1 - \delta] \\ 1, & x^{*,y} \in (1 - \delta, 1]. \end{cases}
$$

The same relation between x^* and y^* can be obtained for the optimal solution in the x direction when $\delta < 0$ and $n \to \infty$. In that case, when we initiate learning in the x direction, we start with $x^{*,x} = -\delta$ for $y^{*,x} = 0$ and continue to follow the relationship $y^{*,x} = \delta + x^{*,x}$ for $x^{*,x} \in (0, 1 + \delta)$. We demonstrate the

effect of the relative strategic advantage on the optimal learning path, by considering the game with the fitness matrix R_3 and the incompetence matrix S_3, referred to as Example 3:

$$R_3 = \begin{pmatrix} 0 & a \\ 1 & 0 \end{pmatrix}, \quad S_3 = \begin{pmatrix} 0.9 & 0.1 \\ 0.9 & 0.1 \end{pmatrix}.$$

Strategy 2 obtains a learning advantage ($\beta_3 = 0.1 \ll \alpha_3 = 0.9$). We give strategy 1 a fitness advantage by selecting three values for a: 5, 7 and 9, which result in $\delta_3 \approx 0.74, 0.28$ and 0, respectively. Then, smaller values of a result in the larger first step (see Figure 2B).

The next natural question to ask would be what is the influence of the number of steps we make. Computing the exact form of Φ_x^* and Φ_y^* at the optimal solutions and taking their rate of change in n yields that

$$\Delta \Phi_x^* = \frac{2(\delta + 1)^3}{3(2n - 3)} \quad \text{and} \quad \Delta \Phi_y^* = \frac{2(-\delta + 1)^3}{3(2n - 3)},$$

which are positive for any $n \geq 2$. For $\delta \ll 0$ we can observe negative Φ_x^* identifying that the x direction is no longer preferred. This indicates that the bigger the difference between skills, the lower potential fitness-over-learning that can be gained. In this sense, the skill with higher incompetence might reduce the fitness as it requires some investments for the skills to be learnt.

Our learning scheme allows for an adjustment of individuals' behaviour in case of any disruption leading to behavioural mistakes. The number of steps required for such an adjustment can be as low as 2 allowing for a quick reaction to system's uncertainty. Moreover, such an adjustment does not require for the system to stop interactions for learning. Individuals can continue interacting with their group-mates while their behavioural mistakes are reduced and fitness is maximised.

6. Conclusions

In this paper, we considered the evolutionary game where two skills coexist in a mixed equilibrium, and hence are both required. This is a key assumption as we aimed to answer the question: If both strategies are important, then how do we learn them in an optimal way? We introduced a fitness-over-learning function which measures the improvement in fitness of the population over the learning path that was taken. This function relies on both performance of the strategy and its rate of mistakes.

The naive suggestion would be that the most advantageous skill in terms of fitness has to be learnt first. However, the strategy with lower relative strategic advantage is learnt first in the optimal learning path. We conjecture that this adjusts the difference between the skills and then, once they are comparable, optimal learning suggests to learn both skills with equal rates. These findings indicate that once disrupted, selection tries to recover the most affected strategies first even if their fitness is not the highest. Nonetheless, if the fitness difference is high enough to overcome the effect of incompetence, then the optimal learning will demand that the better strategy is learned first. Another possible interpretation would be to consider the mixed equilibrium as mixed strategies used by players. Then, by learning the less-advantageous strategy, individuals are reaching the nearest optimal mixed strategy.

Importantly, we parametrised the notion of strategic advantage of cooperation versus defection with a single quantity denoted by δ, which captures the tradeoffs between fitness and propensity to make execution errors in these two modes of behaviour. Interestingly, we showed that this quantity has a critical threshold absolute value of 1. Namely, if $|\delta| < 1$ then our Corollary 1 implies that learning by many small steps is preferable to learning by fewer large steps. Arguably, this captures the belief that complex skills are best learned incrementally. However, if $|\delta| > 1$, then Corollary 1 shows that coexistence is preserved by

one of only two possible learning paths: (a) full learning first in the *x* direction, followed by full learning in the y direction; or (b) the other way around. This suggests that a sufficiently large strategic advantage of cooperation over defection (or the converse) eliminates the luxury of incremental learning.

In addition, the number of steps in the learning path maximising the fitness is not bounded. Indeed, taking many small learning steps improves the fitness we observe. However, as demonstrated in Figure 3, there may exist a number of steps n^* after which the increase in the objective function seems insignificant. Hence, we can determine a sufficient number of steps to achieve a target level of the fitness-over-learning function in applications.

Overall, the learning scheme proposed in this paper can be used to correct behavioural uncertainty when the system already reached its equilibrium but was disrupted. However, our formulation has its limitations. The main limitation is that we allow for only one skill to be learnt at a time. However, restrictiveness of this assumption decreases with increasing number of steps. The second limitation of our scheme is that the direction of the learning path can only be chosen at the very beginning and cannot be changed while individuals are learning. While it may be more natural to permit the learning direction to change at each step, it would also require more resources to be spent on learning. Therefore, the cost of learning would need to be taken into account. Such extensions can be studied in future research.

Author Contributions: Conceptualisation, M.K., S.S.S. and J.A.F.; validation, M.K., S.S.S., J.A.F. and K.C.; formal analysis, M.K. and S.S.S.; investigation, M.K. and S.S.S.; resources, K.C.; writing—original draft preparation, M.K.; writing—review and editing, S.S.S., J.A.F. and K.C.; visualisation, M.K. and S.S.S.; supervision, J.A.F. and K.C.; project administration, M.K.; funding acquisition, M.K., J.A.F. and K.C. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Mathematical Appendix

Here, we provide all the formal proofs for the results in the manuscript.

Lemma A1. *Let $\tilde{\phi}(X)$ be the fitness function and let $C_n^{*x} \in \mathcal{L}_n$. Then Φ_x and Φ_y can be simplified to*

$$\Phi_x(C_n) = L + \sum_{i=1}^{n-1} \left\{ \tilde{a} y_i(\Delta x_i) + \tilde{b} x_{i+1}(\Delta y_i) + \frac{1}{2} y_i(\Delta x_i^2) + \frac{1}{2} x_{i+1}(\Delta y_i^2) \right\} \tag{A1}$$

and

$$\Phi_y(C_n) = L + \sum_{i=1}^{n-1} \left\{ \tilde{a} y_{i+1}(\Delta x_i) + \tilde{b} x_i(\Delta y_i) + \frac{1}{2} y_{i+1}(\Delta x_i^2) + \frac{1}{2} x_i(\Delta y_i^2) \right\}, \tag{A2}$$

where $L := 2\tilde{a}\tilde{b} + \frac{1}{2}(\tilde{a} + \tilde{b})$.

Appendix A.1. Proof of Theorem 1

Proof. Let $\tilde{a} = \tilde{b}$ and let $\{x_k^{*x}, y_k^{*x}\}_{k=1}^{n} \in \mathcal{L}_n$ be the optimal strategy in the x direction. Since

$$\sum_{i=1}^{n-1} [y_i(\Delta x_i) + x_{i+1}(\Delta y_i)] = \sum_{i=1}^{n-1} [-y_i x_i + x_{i+1} y_{i+1}] = \sum_{i=1}^{n-1} \Delta(y_i x_i) = y_n x_n - x_1 y_1 = 1.$$

we obtain in addition to Lemma A1

$$\Phi_x(x_i, y_i) = 2\tilde{a}^2 + 2\tilde{a} + \frac{1}{2} \sum_{i=1}^{n-1} \left[y_i(\Delta(x_i)^2) + x_{i+1}(\Delta(y_i)^2) \right]$$

$$= 2\tilde{a}^2 + 2\tilde{a} + \frac{1}{2} \sum_{i=1}^{n-1} \left[x_{i+1}(\Delta(y_i)^2) + y_i(\Delta(x_i)^2) \right] = \Phi_y(y_i, x_i).$$

Hence, if $\mathcal{C}_n^{*x} = \{(x_i^{*x}, y_i^{*x}) \, 1 \leq i \leq n\}$ maximises Φ_x, then $\mathcal{C}_n^{*y} = \{(y_i^{*x}, x_i^{*x}) \, 1 \leq i \leq n\}$ maximises Φ_y. This also holds in the opposite direction and completes the proof. □

Appendix A.2. Proof of Proposition 1

Proof. If $\mathcal{C}_n^*$ optimises Φ_x, then it necessarily solves

$$\frac{\partial}{\partial x_k} \Phi_x(\mathcal{C}_n^*) = -x_k^*(\Delta y_{k-1}^*) + \frac{1}{2}(\Delta y_{k-1}^*)(y_k^* + y_{k-1}^*) = 0,$$

$$\frac{\partial}{\partial y_k} \Phi_x(\mathcal{C}_n^*) = \frac{1}{2}(\Delta x_k^*)(x_k^* + x_{k+1}^*) - (\Delta x_k^*)y_k^* = 0$$

for $k = 2, \ldots, n - 1$. By conditions (a)–(d), $\Delta x_k^*, \Delta y_k^* \neq 0$, requiring

$$2x_k^* = y_k^* + y_{k-1}^* \tag{A3}$$
$$2y_k^* = x_k^* + x_{k+1}^* \tag{A4}$$

for $k = 2, \ldots, n - 1$. Consequently, for $2 \leq k \leq n - 2$,

$$2x_{k+1}^* \overset{(A3)}{=} y_k^* + y_{k+1}^* \overset{(A4)}{=} \frac{1}{2}(x_k^* + x_{k+1}^*) + \frac{1}{2}(x_{k+1}^* + x_{k+2}^*),$$

that is,

$$\Delta^2 x_k^* = x_{k+2}^* - 2x_{k+1}^* + x_k^* = 0.$$

Therefore, Δx_k^* is constant for $k = 2, \ldots, n - 1$, implying $\Delta x_k^* = \frac{1 - x_2^*}{n-2} =: C$ for $k = 2, \ldots, n - 1$. Hence,

$$x_k^* = x_2^* + (k-2)C = \frac{(k-2) + x_2^*(n-k)}{n-2}, \quad k = 2, \ldots, n.$$

Similarly, we obtain $\Delta^2 y_k^* = 0$ for $k = 1, \ldots, n - 3$ and therefore $\Delta y_k^* = \frac{y_{n-1}^*}{n-2} =: D$ for $k = 1, \ldots, n - 2$. This implies that

$$y_k^* = y_{n-1}^* - (n - 1 - k)D = \frac{y_{n-1}^*(k-1)}{n-2}, \quad k = 1, \ldots, n - 1.$$

In order to obtain x_2^* and y_{n-1}^*, note that (A3) implies $x_2^* = \frac{1}{2}y_2^*$ and (A4) $y_{n-1}^* = \frac{1}{2}(1 + x_{n-1}^*)$. Using the solutions, we have

$$x_2^* = \frac{1}{2}y_2^* = \frac{y_{n-1}^*}{2(n-2)}$$

and therefore

$$y_{n-1}^* = \frac{1}{2}(1 + x_{n-1}^*) = \frac{1}{2}\left(\frac{2n-5}{n-2} + \frac{y_{n-1}^*}{2(n-2)^2}\right).$$

Solving for y_{n-1}^* leads to

$$y_{n-1}^* = \frac{2(n-2)}{2n-3},$$

which implies

$$x_2^* = \frac{y_{n-1}^*}{2(n-2)} = \frac{1}{2n-3}.$$

Clearly $0 < x_2^* < 1$ for $n \geq 3$ by the same argument as $0 < y_{n-1}^* < 1$. We therefore obtain a unique increasing sequence in x^* and y^* that satisfies the necessary condition to maximise Φ_x given by

$$x_k^* = \begin{cases} 0 & k = 1 \\ \frac{2k-3}{2n-3} & k = 2, \ldots, n-1 \\ 1 & k = n \end{cases}, \quad y_k^* = \begin{cases} 0 & k = 1 \\ \frac{2k-2}{2n-3} & k = 2, \ldots, n-1 \\ 1 & k = n \end{cases}.$$

In order to check whether the critical point is a maximum or a minimum, we construct the Hessian matrix, which has the block form

$$H = \begin{bmatrix} D_1 & B \\ B^T & D_2 \end{bmatrix},$$

then its determinant is given by

$$\det(H) = \det(D_1 D_2 - BB^T), \tag{A5}$$

where, at the equilibrium, we have

$$D_1 = \operatorname{diag}(-\Delta y_1, \ldots, -\Delta y_{n-2}) = -cI,$$
$$D_2 = \operatorname{diag}(-\Delta x_2, \ldots, -\Delta x_{n-1}) = -cI,$$

where $c = \frac{2}{2n-3}$ and B is a bi-diagonal matrix with diagonal entries given by $(y_2 - x_2, y_3 - x_3, \ldots, y_{n-1} - x_{n-1})$ and subdiagonal entries given by $(x_3 - y_2, x_4 - y_3, \ldots, x_{n-1} - y_{n-2})$. Then,

$$\det(H) = \det(c^2 I - BB^T). \tag{A6}$$

The matrix $D_1 D_2 - BB^T$ is of the same structure as BB^T (Tridiagonal/Band matrix). The $n-2$ eigenvalues of B are given by its diagonal entries

$$v_B^i = y_i - x_i = 1/(2n-3).$$

Hence, B has one eigenvalue $v_B = 1/(2n-3)$ with multiplicity $n-2$. Applying the singular value decomposition of B, we obtain eigenvalues of BB^T, which is $v_{BB^T} = 1/(2n-3)^2$ with multiplicity $n-2$.

Let v_H be an eigenvalue of H, then

$$0 = \det(H - \nu_H I_{2(n-2)}) = \det\left(\begin{bmatrix} D_1 - \nu_H I_{n-2} & B \\ B^T & D_2 - \nu_H I_{n-2} \end{bmatrix}\right)$$

$$\overset{(A6)}{=} \det((cI_{n-2} - \nu_H I_{n-2})^2 - BB^T)$$

$$= \det(\underbrace{(c - \nu_H)^2}_{\psi} I_{n-2} - BB^T),$$

and therefore ψ is an eigenvalues of BB^T. Hence, the eigenvalues of H are given by $\nu_H = -c \pm \sqrt{\nu_{BB^T}}$ and we obtain

$$\nu_H = -\frac{2}{2n-3} \pm \frac{1}{2n-3} \leq -\frac{2}{2n-3} + \frac{1}{2n-3} < 0.$$

This implies that H evaluated at the proposed optimal path is negative definite. Then, the path $C_n^* = \{x_i^{*x}, y_i^{*x} : 1 \leq i \leq n\}$ maximises Φ_x. By Theorem 1, $\{x_i^{*y}, y_i^{*y}\}_{i=1}^n = \{y_i^{*x}, x_i^{*x}\}_{i=1}^n$ is the optimal solution in the y direction, which completes the proof. $\square$

Appendix A.3. Proof of Proposition 2

Proof. As before, we maximise Φ_x over the learning path. Note that further we shall omit the super-script ^{x}. Taking the partial derivatives yields for $2 \leq k \leq n-1$

$$\frac{\partial}{\partial x_k} \Phi_x \overset{Lem\ 1}{=} (\Delta y_{k-1})\left[-\delta - x_k + \frac{1}{2}(y_{k-1} + y_k)\right] = 0,$$

$$\frac{\partial}{\partial y_k} \Phi_x \overset{Lem\ 1}{=} (\Delta x_k)\left[\delta + \frac{1}{2}(x_k + x_{k+1}) - y_k\right] = 0.$$

By properties (a)–(b), that is, $\Delta x_k \neq 0$ and $\Delta y_k \neq 0, \forall k = 1, \ldots, n-1$, we obtain the necessary conditions for the critical points

$$x_k^* = -\delta + \frac{1}{2}(y_k^* + y_{k-1}^*), \quad 2 \leq k \leq n-1 \tag{A7}$$

$$y_k^* = \delta + \frac{1}{2}(x_k^* + x_{k+1}^*), \quad 2 \leq k \leq n-1. \tag{A8}$$

Together, we obtain for $3 \leq k \leq n-1$

$$x_k^* \overset{(A7)}{=} -\delta + \frac{1}{2}(y_k^* + y_{k-1}^*) \overset{(A8)}{=} \frac{1}{2}x_k^* + \frac{1}{4}x_{k+1}^* + \frac{1}{4}x_{k-1}^*$$

or simply

$$0 = -2x_k^* + x_{k+1}^* + x_{k-1}^* = \Delta^2 x_{k-1}^*, \quad 3 \leq k \leq n-1.$$

This implies that $\Delta x_k^* := c \in \mathbb{R}$ for $2 \leq k \leq n-1$. Then, the interval $[x_2^*, 1]$ is divided into $n-2$ equal sized segments. Consequently, $c = \frac{1 - x_2^*}{n-2}$ and therefore for $2 \leq k \leq n$

$$x_k^* = x_2^* + (k-2)c = x_2^* + (k-2)\frac{1 - x_2^*}{n-2} = \frac{n-k}{n-2}x_2^* + \frac{k-2}{n-2}. \tag{A9}$$

Similarly, we obtain $\Delta^2 y_k^* = 0, \forall k = 1, \ldots, n-3$. Hence, $\Delta y_k^* := d \in \mathbb{R}$ for $1 \leq k \leq n-2$, implying that the interval $[0, y_{n-1}^*]$ is distributed into $n-2$ equally sized segments, such that $d = \frac{y_{n-1}^*}{n-2}$. Then,

$$y_k^* = (k-1)d = (k-1)\frac{y_{n-1}^*}{n-2}, \quad 1 \le k \le n-1. \tag{A10}$$

To find x_2^* and y_{n-1}^*, we utilise relations (A7) and (A8) as follows

$$x_2^* \overset{(A7)}{=} -\delta + \frac{1}{2}y_2^* \overset{(A10)}{=} -\delta + \frac{y_{n-1}^*}{2(n-2)}$$

and

$$y_{n-1}^* \overset{(A8)}{=} \delta + \frac{1}{2}x_{n-1}^* + \frac{1}{2} \overset{(A9)}{=} \delta + \frac{1}{2} + \frac{1}{2}\left(x_2^* + (n-3)\frac{1-x_2^*}{n-2}\right).$$

Combining both equations yields

$$x_2^* = -\delta + \frac{1+\delta}{2n-3} \quad \text{and} \quad y_{n-1}^* = (1+\delta)\frac{2(n-2)}{2n-3},$$

which is only feasible for $\delta < \frac{1}{2(n-2)}$ to guarantee $x_2^*, y_{n-1}^* \in (0,1]$, or if $\delta < -1$, then it is only feasible for $n = 2$. Hence, the only path satisfying the necessary condition to be the optimal path in the x direction is given by

$$x_k^* = \begin{cases} 0 & k = 1 \\ -\delta + (1+\delta)\frac{2k-3}{2n-3} & 2 \le k \le n \end{cases}, \quad y_k^* = \begin{cases} (1+\delta)\frac{2k-2}{2n-3} & 1 \le k \le n-1 \\ 1 & k = n \end{cases}.$$

To support the claim that this learning path maximises the fitness-over-learning function we consider the Hessian matrix for Φ_x

$$H = \begin{bmatrix} D_1 & B \\ B^T & D_2 \end{bmatrix},$$

where D_1 and D_2 are diagonal matrices of the form

$$D_1 = -\text{diag}(\Delta y_1, \Delta y_2, \dots, \Delta y_{n-2}),$$

$$D_2 = -\text{diag}(\Delta x_2, \Delta x_3, \dots, \Delta x_{n-1}),$$

and B being a bi-diagonal matrix with diagonal entries given by

$$(-\delta + (y_2 - x_2), -\delta + (y_3 - x_3), \dots, -\delta + (y_{n-1} - x_{n-1})),$$

and subdiagonal entries

$$(-\delta + (x_3 - y_2), -\delta + (x_4 - y_3), \dots, -\delta + (x_{n-1} - y_{n-2})).$$

As in Proposition 1 we obtain

$$\det(H) = \det(D_1 D_2 - BB^T)$$

and the eigenvalues of H, ν_H, satisfy

$$(c + \nu_H)^2 = \nu_{BB^T} = (\nu_B)^2 \implies \nu_H = -c \pm |\nu_B|,$$

where ν_B are the eigenvalues of B given by the diagonal entries, namely, $\nu_B^i = -\delta + (y_i^* - x_i^*)$ and $c = \Delta x_i^* = \Delta y_{i-1}^* = \frac{2(1+\delta)}{(2n-3)}$. For a feasible solution, $c > 0$. Hence the eigenvalues of H are negative as

$$| -\delta + (y_i^* - x_i^*)| = \frac{|1 + \delta|}{2n - 3} < 2\frac{|1 + \delta|}{(2n - 3)}.$$

Therefore, $|v_B| < c$ and $v_H < 0$. This completes the proof that the learning path is optimal for the x direction. $\square$

Appendix A.4. Proof of Proposition 3

Proof. The partial derivatives of Φ_y are

$$\frac{\partial}{\partial y_k}\Phi_y \overset{\text{Lem 1}}{=} \frac{\partial}{\partial y_k}\left\{2\tilde{a}\tilde{b} + \frac{1}{2}(\tilde{a} + \tilde{b}) + \sum_{i=1}^{n-1}\left\{\tilde{a}y_{i+1}(\Delta x_i) + \tilde{b}x_i(\Delta y_i) + \frac{1}{2}y_{i+1}(\Delta x_i^2) + \frac{1}{2}x_i(\Delta y_i^2)\right\}\right\}$$

$$= \tilde{a}(\Delta x_{k-1}) + \tilde{b}x_{k-1} - \tilde{b}x_k + \frac{1}{2}(\Delta x_{k-1}^2) + \frac{1}{2}x_{k-1}2y_k - \frac{1}{2}x_k 2y_k$$

$$= (\Delta x_{k-1})\left[\delta + \frac{1}{2}(x_k + x_{k-1}) - y_k\right],$$

$$\frac{\partial}{\partial x_k}\Phi_y \overset{\text{Lem 1}}{=} \frac{\partial}{\partial x_k}\left\{2\tilde{a}\tilde{b} + \frac{1}{2}(\tilde{a} + \tilde{b}) + \sum_{i=1}^{n-1}\left\{\tilde{a}y_{i+1}(\Delta x_i) + \tilde{b}x_i(\Delta y_i) + \frac{1}{2}y_{i+1}(\Delta x_i^2) + \frac{1}{2}x_i(\Delta y_i^2)\right\}\right\}$$

$$= \tilde{a}y_k - \tilde{a}y_{k+1} + \tilde{b}(\Delta y_k) + \frac{1}{2}y_k 2x_k - \frac{1}{2}y_{k+1}2x_k + \frac{1}{2}(\Delta y_k^2)$$

$$= (\Delta y_k)\left[-\delta - x_k + \frac{1}{2}(y_{k+1} + y_k)\right].$$

By properties (a)–(b), that is, $\Delta x_k, \Delta y_k \neq 0$ for $k = 1, \ldots, n - 1$, we obtain the necessary conditions

$$\delta + \frac{1}{2}(x_k + x_{k-1}) - y_k = 0, \qquad 2 \leq k \leq n - 1$$

$$-\delta - x_k + \frac{1}{2}(y_{k+1} + y_k) = 0, \qquad 2 \leq k \leq n - 1,$$

which can be rewritten as

$$x_k^* = -\delta + \frac{1}{2}(y_{k+1} + y_k), \qquad 2 \leq k \leq n - 1 \tag{A11}$$

$$y_k^* = \delta + \frac{1}{2}(x_k + x_{k-1}), \qquad 2 \leq k \leq n - 1. \tag{A12}$$

In the similar manner as in the proof of Proposition 2, we obtain

$$\Delta y_k^* = \frac{1 - y_2^*}{n - 2} \text{ for } 2 \leq k \leq n - 1 \quad \text{and} \quad \Delta x_k^* = \frac{x_{n-1}^*}{n - 2} \text{ for } 1 \leq k \leq n - 2, \tag{A13}$$

and finally the optimal solution of the form

$$y_k^* = \begin{cases} 0, & k = 1 \\ \delta + (1 - \delta)\frac{2k-3}{2n-3}, & k = 2, \ldots, n - 1 \\ 1, & k = n \end{cases}, \quad x_k^* = \begin{cases} 0, & k = 1 \\ (1 - \delta)\frac{2k-2}{2n-3}, & k = 2, \ldots, n - 1 \\ 1, & k = n \end{cases}.$$

This is feasible only if $\delta > -\frac{1}{2(n-2)}$ to guarantee $y_k^*, x_k^* \geq 0$ for $1 \leq k \leq n$. In the similar manner to Proposition 2 one can show that it is, indeed, a maximum of Φ_y. $\square$

Appendix A.5. Proof of Theorem 2

Proof. We use the fitness-over-learning functions at the equilibrium and obtain

$$\Phi_x^* - \Phi_y^* = -\delta \times \frac{3\delta^2(n-1)(n-2)+1}{(2n-3)^2}.$$

Since $\dfrac{3\delta^2(n-1)(n-2)+1}{(2n-3)^2} > 0$, for $\delta > 0$ the y direction is optimal and for $\delta < 0$ it is the x direction. □

Appendix A.6. Proof of Corollary 1

Proof. The statement follows from the fact that for $\delta \geq 1$, the optimal learning path of order n in the y direction is only feasible for $n = 2$, resulting in a one-step function. Since the optimal path in the x direction is unfeasible for such δ for all n, the y direction is chosen first. □

Appendix A.7. Proof of Corollary 2

Proof. The optimal solution in the y direction satisfies the condition:

$$y_k^{*,y} = \delta + \frac{(1-\delta)(2k-3)}{2n-3} = \delta + \frac{(1-\delta)(2k-2)}{2n-3} - \frac{(1-\delta)}{2n-3} = \delta + x_k^{*,y} - \frac{1-\delta}{2n-3}.$$

Hence,

$$y^{*,y} := \lim_{n\to\infty} y_k^{*,y} = \lim_{n\to\infty}(\delta + x_k^{*,y}),$$

or, since as $n \to \infty$, then $k \to \infty$, we simply obtain

$$y^{*,y} = \delta + x^{*,y}, \ \forall x^{*,y} \in (0, 1-\delta).$$

Boundary values for $x^{*,y} = 0$ and $x^{*,y} \in (1-\delta, 1]$ follow from the constraints $0 \leq y^{*,y} \leq 1$. □

References

1. Broom, M.; Jan, R. *Game-Theoretical Models in Biology*; CRC Press, Taylor & Francis Group: Boca Raton, FL, USA, 2013.
2. Weibull, J.W. *Evolutionary Game Theory*; MIT Press: Cambridge, MA, USA, 1997.
3. Fudenberg, D.; Levine, D.K. *The Theory of Learning in Games*; The MIT Press: Cambridge, MA, USA, 1999.
4. Fudenberg, D.; Levine, D. Learning in games. *Eur. Econ. Rev.* **1998**, *42*, 631–639. [CrossRef]
5. Selten, R. Reexamination of the perfectness concept for equilibrium points in extensive games. *Int. J. Game Theory* **1975**, *4*, 25–55. [CrossRef]
6. Stadler, P.F.; Schuster, P. Mutation in Autocatalytic Reaction Networks. *J. Math. Biol.* **1992**, *30*, 597–631. [CrossRef] [PubMed]
7. Tarnita, C.E.; Antal, T.; Nowak, M.A. Mutation–selection equilibrium in games with mixed strategies. *J. Theor. Biol.* **2009**, *261*, 50–57. [CrossRef] [PubMed]
8. Wu, B.; Gokhale, C.S.; Wang, L.; Traulsen, A. How small are small mutation rates? *J. Math. Biol.* **2012**, *64*, 803–827. [CrossRef] [PubMed]
9. Archibald, C.; Shoham, Y. Hustling in repeated zero-sum games with imperfect execution. In Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Spain, 16–22 July 2011.
10. Bowling, M.; Veloso, M. Existence of multiagent equilibria with limited agents. *J. Artif. Intell. Res.* **2004**, *22*, 353–384. [CrossRef]

11. Borkar, V.S.; Jain, S.; Rangarajan, G. Evolutionary games with two timescales. *Phys. D Nonlinear Phenom.* **1999**, *125*, 155–166. [CrossRef]
12. Beck, J.D.; Ejov, V.; Filar, J.A. Incompetence and impact of training in bimatrix games. *Automatica* **2012**, *48*, 2400–2408. [CrossRef]
13. Beck, J.D. Incompetence, Training and Changing Capabilities in Game Theory. Ph.D. Thesis, University of South Australia, Adelaide, Australia, 2013.
14. Kleshnina, M.; Filar, J.A.; Ejov, V.; McKerral, J.C. Evolutionary games under incompetence. *J. Math. Biol.* **2018**, *77*, 627–646. [CrossRef]
15. Smith, J.M. *Evolution and the Theory of Games*; Cambridge University Press: Cambridge, UK, 1982.
16. May, R.M. More evolution of cooperation. *Nature* **1987**, *327*, 15–17. [CrossRef]
17. Sigmund, K.; Silva, H.D.; Hauert, C.; Traulsen, A. Social learning promotes institutions for governing the commons. *Nature* **2010**, *466*, 861. [CrossRef]
18. Nowak, M.A. *Evolutionary Dynamics: Exploring the Equations of Life*; The Belknap Press of Harvard University Press: Cambridge, MA, USA, 2006.
19. Ågren, J.A.; Davies, N.G.; Foster, K.R. Enforcement is central to the evolution of cooperation. *Nat. Ecol. Evol.* **2019**, *3*, 1018–1029. [CrossRef] [PubMed]
20. Axelrod, R. The emergence of cooperation among egoists. *Am. Political Sci. Rev.* **1981**, *75*, 306–318. [CrossRef]
21. Axelrod, R.; Hamilton, W.D. The evolution of cooperation. *Science* **1981**, *211*, 1390–1396. [CrossRef]
22. Doebeli, M.; Hauert, C. Models of cooperation based on the Prisoner's Dilemma and the Snowdrift game. *Ecol. Lett.* **2005**, *8*, 748–766. [CrossRef]
23. Chao, L.; Elena, S.F. Nonlinear trade-offs allow the cooperation game to evolve from Prisoner's Dilemma to Snowdrift. *Proc. R. Soc. B Biol. Sci.* **2017**, *284*, 20170228. [CrossRef]
24. Anh, H.T.; Pereira, L.M.; Santos, F.C. Intention recognition promotes the emergence of cooperation. *Adapt. Behav.* **2011**, *19*, 264–279.
25. Segbroeck, S.V.; Jong, S.D.; Nowé, A.; Santos, F.C.; Lenaerts, T. Learning to coordinate in complex networks. *Adapt. Behav.* **2010**, *18*, 416–427. [CrossRef]
26. Hilbe, C.; Schmid, L.; Tkadlec, J.; Chatterjee, K.; Nowak, M.A. Indirect reciprocity with private, noisy, and incomplete information. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 12241–12246. [CrossRef]
27. Draghi, J.; Whitlock, M.C. Overdominance interacts with linkage to determine the rate of adaptation to a new optimum. *J. Evol. Biol.* **2015**, *28*, 95–104. [CrossRef] [PubMed]
28. Hofbauer, J.; Sigmund, K. Evolutionary game dynamics. *Bull. Am. Math. Soc.* **2003**, *40*, 479–519. [CrossRef]
29. Taylor, P.D.; Jonker, L.B. Evolutionary stable strategies and Game Dynamics. *Math. Biosci.* **1978**, *40*, 145–156. [CrossRef]
30. Hofbauer, J.; Schuster, P.; Sigmund, K.; Wolff, R. Dynamical Systems Under Constant Organization II: Homogeneous Growth Functions of Degree p = 2. *SIAM J. Appl. Math.* **1980**, *38*, 282–304. [CrossRef]
31. Kleshnina, M. Evolutionary Games under Incompetence & Foraging Strategies of Marine Bacteria. Ph.D. Thesis, University of Queensland, Brisbane, Australia, 2019.
32. Cressman, R.; Tao, Y. The replicator equation and other game dynamics. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 10810–10817. [CrossRef]