

Supplementary Information

Introspection dynamics in asymmetric multiplayer games

Marta Couto¹ and Saptarshi Pal¹

¹Max Planck Research Group Dynamics of Social Behavior, Max Planck Institute for Evolutionary Biology, 24306 Ploen, Germany

Abstract

The simplest form of social interaction occurs between two nearly similar individuals. This scenario can be stylized in a pairwise symmetric game. More complex behavior emerges when instead we consider multiple interacting individuals, and they may differ significantly among themselves. This case can be captured by a multiplayer asymmetric game. The complexity of these games increases both with the number and diversity of players. From evolutionary dynamics to learning models, game theoretic frameworks have been widely used to study strategic behavior. Introspection dynamics has proven to be a useful learning model for tackling complex games. It provides a simple way to compute the average abundance of each combination of strategies (or actions), that is, the stationary distribution of strategies. Here, our aim is to analyse multiplayer and (a)symmetric games. We extend previous analytical results on pairwise games under introspection dynamics to multiplayer games. For general games, we obtain the analytical stationary distribution for 2 strategies, symmetric games. Moreover, we consider a particular class of games – additive games. In additive games, for any player, the difference between the payoffs provided by any two actions is constant, regardless of the co-players' actions. For these games, we derive the stationary distribution analytically, for any number of players and strategies, with no restriction to symmetric games. To illustrate our theoretical results, we analyse various multiplayer asymmetric social dilemmas. In general, players that have a lower cost or a higher benefit of cooperation learn to cooperate more frequently.

Introduction

The simplest form of social interaction occurs between two individuals. These pairwise interactions have been studied extensively[?]. Despite their simplicity, they provide important insights, e.g., on how social species can maintain cooperation[?]. An ubiquitous example of a model game for two players is the prisoner's dilemma. There, the two individuals have a temptation to cheat on each other, while both would be better off by offering mutual help. Solving this apparent paradox has been the focus of many studies, where various explanations were proposed[?].

Yet, many interesting collective behavior occur when multiple individuals interact simultaneously¹⁻⁸. Notably, most of these situations cannot be captured by the sum of several pairwise interactions. Thus, to account for such non-linearities we need to consider multiplayer games⁶. One well-known effect that only emerges when more than two players are present is the second-order free riding problem⁹. A natural solution to maintain pro-social behavior in a community is to monitor and punish defectors (and/or reward cooperators). However, most forms of sanctioning are considerably costly[?]. Therefore, an additional (second-order) dilemma is at stake: deterring defection is beneficial for all, but an individual that does not pay the associated cost has an advantage over the ones that do. As such, peer-punishment and sanctioning institutions can be hard to sustain without additional mechanisms or a suitable design¹⁰.

Another interesting aspect of multiplayer games is the scale or size of the interaction itself. Are multiplayer social dilemmas harder to solve than small-groups'??². Indeed, in situations that require some sort of coordination and expectations on others play an important role in one's decisions (like in stag-hunt games), it is easy to see how a growing group size might hinder the optimal outcome – it is hard to cooperate in large groups[?]. This is not a general effect, though. It can happen that adding more players actually promotes cooperation. As a biological example, consider a task allocation problem in a bacterial system. A strain of bacteria can do any of three actions: to specialize in producing one of two kinds of enzymes (at some personal cost) or not to produce any enzyme. The bacteria need the two different types of enzymes at the same time to obtain resources from the environment. In¹¹, authors investigate both the 2-player and 3-player cases of this scenario. Although the underlying rules defining the interactions are the same for the two cases, simply introducing a third player changes the overall game, and hence the outcomes. With three players, the production of one of the enzymes increases while freeloading decreases, comparing to the 2-player case. Additionally, not only the average group size can have an important effect, but also the variance of the group size distribution^{12,13}.

Complexity further increases when players differ significantly among themselves. This diversity can be captured by asymmetric games¹⁴⁻²². In symmetric games, all players are indistinguishable. Thus, to fully characterise a state of the game, we only require to know the number of players playing each strategy. Conversely, in asymmetric games, players can be different in several ways. We say they are

of different types or roles, which can differ in payoffs and strategies. Therefore, they can have uneven effects on others' payoffs too. For example, in a public goods game, even if every player chooses to cooperate, their contributions to the common good or, alternatively, the cost they pay might be diverse²². Hence, to fully describe a state of the game, we need to know which of the players is doing what. This greatly increases the size of the game's state space; even more so, for more than two players.

From this, we see how the mathematical analysis of multiplayer asymmetric games can become cumbersome, the extreme cases being many-player games where all individuals are different. Perhaps for that reason, there are not many studies on multiplayer asymmetric games, compared to their simpler counterparts. Some exceptions are²². Still, we believe a general approach is lacking. As such, our aim here is to narrow that gap, by studying multiplayer, (a)symmetric games. In the following, we discuss several canonical frameworks from the literature that have been used to study both multiplayer and asymmetric games.

Related Work

From evolutionary game theory^{23–26} to learning models^{27–30}, game theoretic frameworks have been widely used to study strategic behavior. In evolutionary game theory, we assume a population of players and a mechanism by which strategies evolve over time. More successful strategies tend to persist in the population, either by genetic or cultural inheritance. The first key concept in evolutionary game theory is “evolutionary stable strategy” (ESS), introduced in 1973, in the seminal paper of John Maynard Smith and George Price²³. A monomorphic population playing an ESS cannot be invaded by mutants. The concept of ESS, originally proposed for pairwise encounters, was already extended for multiplayer games in^{1,31,32}. Although suggestive of an evolutionary process, ESS does not really describe a dynamical system. In the more classical infinite population models, ordinary differential equations are the central mathematical tools²⁵. One of the most famous of these dynamics is the replicator equation^{25,33}. Replicator equation has been widely used and easily comprises multiplayer games^{12,34–37}. More recently, also the replicator-mutator equation was applied to study the dynamics of multiplayer games³⁸. As for asymmetric games, a few additional assumptions are needed. For example, if there are two different types of players, typically, either there are two populations co-evolving (“bimatrix games”^{25,39}) or there is a single population of players where each can play the two types or roles (“role games”)²⁵. In the latter kind of model, a strategy must include the action played in each of the two roles.

To account for stochastic effects naturally present in finite populations, another set of tools was developed, known as stochastic evolutionary game dynamics.^{40,41} Besides the finiteness of population sizes, it also considers random mutations. Here, the quantities of interest are fixation probabilities and fixation times (i.e., the probability and average time that a single mutant overtakes an initially monomorphic population) for when mutations are rare[?], or average strategy abundance (i.e., the stationary distribution

of strategy frequencies) for when mutation rates take intermediate values^{42,43}.

These frameworks have been applied to asymmetric games assuming co-evolving (finite) subpopulations of different player types. Fixation probabilities for asymmetric 2-player⁴⁴ and 3-player games[?] were recently derived. Whereas with intermediate mutation, strategy abundances were obtained only for 2-player asymmetric games[?] or multiplayer symmetric games^{11,35,45}. For a review on evolutionary multiplayer games both in infinitely large populations as well as in finite populations, we refer to⁶.

Different from evolutionary game theory, learning models (of strategic behavior) take another approach^{18,27–30,46–49}. There is no evolution of strategies in a population necessarily, but a process by which individuals learn strategies dynamically. Despite of the conceptual disparities, there have been some efforts in unifying the two lines of research of evolutionary game theory and multiagent reinforcement learning^{29,46}.

Clearly, we can distinguish two different pathways that led to “learning in games”: one which stemmed from the broad field of reinforcement learning²⁷ and the other from classical game theory itself[?]. The “theory of learning in games” was proposed as a refinement of classical equilibrium concepts, like Nash equilibrium, where strict assumptions of common knowledge of rationality are made. Instead, learning consists of “*an alternative explanation that equilibrium arises as the long-run outcome of a process in which less than fully rational players grope for optimality over time*”[?]. In this context, several learning models were introduced, such as fictitious play[?].

Introspection dynamics has proven to be a useful learning model for tackling 2-player, asymmetric games⁵⁰. This feature comes from the fact that players update their strategies by exploring their own set of strategies in the following simple way: at each time, after a round of the game, a player considers a random alternative strategy; they compare the payoff that it would have given them to their current payoff; if the new strategy would provide a higher payoff, the player more likely adopts it on the next round. While in⁵⁰ only 2-player games were considered, this framework is general enough to account for multiple players. Indeed, as each player assumes that all others remain the same when doing the counterfactual comparison, introspection dynamics is an appropriate model to deal with the complexity introduced by multiple players.

Here, we extend previous analytical results of pairwise games under introspection dynamics to multiplayer games. For general games, we obtain the analytical stationary distribution for 2-strategies symmetric games. For a particularly simple class of games – additive games – we derive the stationary distribution of strategies analytically, for any number of players and strategies, and with no restriction to symmetric games. To illustrate our theoretical results, we analyse various multiplayer asymmetric social dilemmas, extending the framework in³⁴ to asymmetric games.

Model

We consider the normal form game with N players where $N > 2$. In the game, a player, say player i , can play actions from their action set, $\mathbf{A}_i := \{a_{i,1}, a_{i,2}, \dots, a_{i,m_i}\}$. The action set of player i has m_i actions. There are, therefore, $m_i \times m_2 \times \dots \times m_N$ distinct states of the game. We denote a state of the game by collecting the actions of all the players in the game in a vector, $\mathbf{a} := (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N)$ where $\mathbf{a} \in \mathbf{A} := \mathbf{A}_1 \times \mathbf{A}_2 \times \dots \times \mathbf{A}_N$. We also use the notation, $\mathbf{a} := (\mathbf{a}_i, \mathbf{a}_{-i})$ to denote the state from the perspective of player i . In the state $(\mathbf{a}_i, \mathbf{a}_{-i})$, player i plays the action $\mathbf{a}_i \in \mathbf{A}_i$ and their co-players play the action $\mathbf{a}_{-i} \in \mathbf{A}_{-i} := \prod_{j \neq i} \mathbf{A}_j$. The payoff of a player depends on the state of the game. We denote the payoff of player i in the state \mathbf{a} with $\pi_i(\mathbf{a})$ or $\pi_i(\mathbf{a}_i, \mathbf{a}_{-i})$.

The players update their strategies over time using the introspection dynamics[?]. At every time step, one randomly chosen player can update their strategy. The randomly chosen player, say i , currently playing strategy $a_{i,k}$, compares their current payoff to the payoff that they would obtain if they played a randomly selected strategy, $a_{i,l} \neq a_{i,k}$, from their action set \mathbf{A}_i . This comparison is done while assuming that the co-players do not change their respective plays. When the co-players of player i play \mathbf{a}_{-i} , player i changes to the new strategy $a_{i,l}$ from $a_{i,k}$ with the probability,

$$p_{a_{i,k} \rightarrow a_{i,l}}(\mathbf{a}_{-i}) = \frac{1}{1 + e^{-\beta(\pi_i(a_{i,l}, \mathbf{a}_{-i}) - \pi_i(a_{i,k}, \mathbf{a}_{-i}))}} \quad (1)$$

in the next round. Here $\beta \in [0, \infty)$ is the selection strength parameter that represents the importance that players give to payoff differences while updating their strategies. At $\beta = 0$, players update to a randomly chosen strategy with probability 0.5. For $\beta > 0$, players update to new strategy with probability greater than 0.5 (or less than 0.5) if the switch gives them a non-zero increase (or decrease) in the payoffs. The probability that the chosen player makes no-update is then,

$$p_{a_{i,k} \rightarrow a_{i,k}}(\mathbf{a}_{-i}) = 1 - \sum_{k \neq l} p_{a_{i,k} \rightarrow a_{i,l}}(\mathbf{a}_{-i}) \quad (2)$$

Introspection dynamics can be studied by analyzing properties of the resulting transition matrix, \mathbf{T} . The transition matrix element $\mathbf{T}_{\mathbf{a}, \mathbf{b}}$ denotes the conditional probability that the game goes to the state \mathbf{b} in the next round if it is in state \mathbf{a} in the current round. We define the neighbourhood set of \mathbf{a} as the following:

Definition 1 (Neighbourhood set of a state). *The neighbourhood set of state \mathbf{a} , $\text{Neb}(\mathbf{a})$, is defined as the following:*

$$\text{Neb}(\mathbf{a}) := \{\mathbf{b} \in \mathbf{A} \setminus \{\mathbf{a}\} : (\exists j)[\mathbf{b}_{-j} = \mathbf{a}_{-j} \wedge \mathbf{b}_j \neq \mathbf{a}_j]\} \quad (3)$$

In other words, a state in $\text{Neb}(\mathbf{a})$ is a state that has exactly one player playing a different action than in state \mathbf{a} . Consider the game where there are three players and each player has the identical action set $\{C, D\}$. The state (C, C, C) is in the neighbouring set of (C, C, D) whereas the state (C, D, D) is not in the neighbourhood set of (C, C, C) . Two states that belong in each other's neighbourhood set only differ in exactly a single player's (that we call as the index of difference) action.

Definition 2 (Index of difference between neighbouring states). *If two states, \mathbf{a} and \mathbf{b} , satisfy $\mathbf{a} \in \text{Neb}(\mathbf{b})$, the index of difference between them, $I(\mathbf{a}, \mathbf{b})$, is the unique integer that satisfies:*

$$\mathbf{a}_{I(\mathbf{a}, \mathbf{b})} \neq \mathbf{b}_{I(\mathbf{a}, \mathbf{b})} \quad (4)$$

In the previous example, the index of difference between the neighbouring states (C, C, C) and (C, C, D) is 3. The third player's action is the difference between the two neighbouring states. Using the above definitions, one can formally define the transition matrix of the introspection dynamics with:

$$\mathbf{T}_{\mathbf{a}, \mathbf{b}} = \begin{cases} \frac{1}{N(m_j-1)} \cdot p_{\mathbf{a}_j \rightarrow \mathbf{b}_j}(\mathbf{a}_{-j}) & \text{if } \mathbf{b} \in \text{Neb}(\mathbf{a}) \text{ and, } j = I(\mathbf{a}, \mathbf{b}) \\ 0 & \text{if } \mathbf{b} \notin \text{Neb}(\mathbf{a}) \\ 1 - \sum_{\mathbf{c} \neq \mathbf{b}} \mathbf{T}_{\mathbf{a}, \mathbf{c}} & \text{if } \mathbf{a} = \mathbf{b} \end{cases} \quad (5)$$

The transition matrix is a row stochastic matrix (the sums of the rows are 1). This implies that the stationary distribution of \mathbf{T} : a left eigenvector of \mathbf{T} corresponding to eigenvalue 1, always exists. The following proposition introduces a sufficient condition for this stationary distribution to be unique.

Proposition 1. *When β is finite, the transition matrix of the introspection dynamics has a unique stationary distribution.*

Proof: A finite value of β results in non-zero probability of transition between neighbouring states. Since no state is isolated (i.e., every state belongs in the neighbourhood set of another state), every state is reachable in a finite number of steps from a starting point with non-zero probability. The transition

matrix \mathbf{T} is therefore primitive for a finite β . By the Perron-Frobenius theorem, a primitive matrix, \mathbf{T} will have a unique and strictly positive stationary distribution $\mathbf{u} := (\mathbf{u}_{\mathbf{a}})_{\mathbf{a} \in \mathbf{A}}$ which satisfies the conditions:

$$\mathbf{u}\mathbf{T} = \mathbf{u} \quad (6)$$

$$\mathbf{u}\mathbf{1} = 1 \quad (7)$$

where $\mathbf{1}$ is the column vector with size same as \mathbf{u} and has all elements as 1. The above equations only present an explicit representation of the stationary distribution \mathbf{u} . The stationary distribution can be explicitly calculated by the following expression (which is derived using Eq. (6) and (7) as:

$$\mathbf{u} = \mathbf{1}^\top (\mathbf{1} + \mathbf{U} - \mathbf{T})^{-1} \quad (8)$$

where \mathbf{U} is a square matrix of size same as \mathbf{T} with all elements 1 and $\mathbf{1}$ is the identity matrix. For all the analytical results in this paper, we consider β to be finite so that stationary distribution of the processes are unique. The stationary distribution element $\mathbf{u}_{\mathbf{a}}$ is the probability that state \mathbf{a} will be played by the players in the long run. Using the stationary distribution, one can calculate the marginal probabilities corresponding to each player's actions. For example, the probability that player i plays action $a_{i,k}$ in the long run, $\xi_{i:a_{i,k}}$, can be computed as,

$$\xi_{i:a_{i,k}} := \sum_{\mathbf{q} \in \mathbf{A}_{-i}} \mathbf{u}_{(a_{i,k}, \mathbf{q})} \quad (9)$$

Since the stationary distribution is a probability distribution, marginal distributions also have the same property. That is, for a player i ,

$$\sum_{k=1}^{m_i} \xi_{i:a_{i,k}} = 1 \quad (10)$$

Additive games and their properties under introspection dynamics

In this section we discuss the stationary properties of the introspection dynamics when players learn to play strategies in a special class of games: the additive games^{??}. In an additive game, the payoff difference that a player earns by making a unilateral switch in their actions is independent of what their co-players play. In other words, if none of the co-players change their current actions, the payoff difference earned by making a switch in actions is *only* determined by the switch and not on the actions of the co-players'. Formally, in additive games, for any player i , any pair of actions $x, y \in \mathbf{A}_i$, and any $\mathbf{q} \in \mathbf{A}_{-i}$,

$$\pi_i(x, \mathbf{q}) - \pi_i(y, \mathbf{q}) =: f_i(x, y) \quad (11)$$

is independent of \mathbf{q} and only dependent on x and y . In the literature, this property is sometimes called the property of *equal gains from switching*[?]. For games with this property, the stationary distribution of

introspection dynamics takes a simple form;

Proposition 2. *When β is finite, the unique stationary distribution, $\mathbf{u} = (\mathbf{u}_{\mathbf{a}})_{\mathbf{a} \in \mathbf{A}}$, of the introspection dynamics for the N -player additive game is given by:*

$$\mathbf{u}_{\mathbf{a}} = \prod_{j=1}^N \frac{1}{\sum_{a' \in \mathbf{A}_j} e^{\beta f_j(a', \mathbf{a}_j)}} \quad (12)$$

where, $f_j(a', \mathbf{a}_j) = \pi_j(a', \mathbf{q}) - \pi_j(\mathbf{a}_j, \mathbf{q})$ is the co-player independent payoff difference that j earns in an additive game when they unilaterally switch their play from \mathbf{a}_j to a' .

For any finite value of the selection strength parameter, β , the stationary distribution of the process for an additive game can be exactly computed using the expression in Eq. (12). For a proof, see Appendix. Using the stationary distribution, one can also exactly compute the cumulative probabilities with which players play their actions in the long run (i.e., the marginal distribution) using Eq. (9). In this regard, introspection learning in additive games is particularly interesting. The stationary distribution and the marginal distributions of introspection learning in additive games are related in a special way,

Proposition 3. *Let $\mathbf{u} = (\mathbf{u}_{\mathbf{a}})_{\mathbf{a} \in \mathbf{A}}$ be the unique stationary distribution of the introspection dynamics with finite β for the N -player additive game. Then, $\mathbf{u}_{\mathbf{a}}$ is the product of the marginal probabilities that each player plays their respective actions in \mathbf{a} . That is,*

$$\mathbf{u}_{\mathbf{a}} = \prod_{j=1}^N \xi_{j:\mathbf{a}_j} \quad (13)$$

where the marginal probability, $\xi_{j:\mathbf{a}_j}$, is the cumulative probability that player j plays \mathbf{a}_j at the stationary distribution. The marginal probability for the N -player additive game is given by:

$$\xi_{j:\mathbf{a}_j} = \frac{1}{\sum_{a' \in \mathbf{A}_j} e^{\beta f_j(a', \mathbf{a}_j)}} \quad (14)$$

where, $f_j(a', \mathbf{a}_j) = \pi_j(a', \mathbf{q}) - \pi_j(\mathbf{a}_j, \mathbf{q})$ is the co-player independent payoff difference that j earns in an additive game when they unilaterally switch their play from \mathbf{a}_j to a' .

The above proposition states that for additive games, the stationary distribution of the introspection dynamics can be factorized into its corresponding marginals. In the long run, the probability that players simultaneously play the state $(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N)$ is the product of the cumulative probabilities that player 1 plays \mathbf{a}_1 , player 2 plays \mathbf{a}_2 and so on. While this factorizing property of the stationary distribution appears for additive games, we do not show that this property is unique to additive games. We use the well-studied example of the linear public goods game to demonstrate the stationary properties of the

introspection dynamics in additive games.

Example of an additive game: linear public goods game with 2 actions

In the simplest version of the linear public goods game with N —players, each player has two possible actions, to contribute (action, C, to cooperate), or to not contribute (action, D, to defect) to the public good. The players differ in their cost of cooperation and the benefit they provide by contributing to the public good. We denote the cost of cooperation for player i and the benefit that they provide to the public good with c_i and b_i respectively. In addition, we define a mapping function $\alpha(\cdot)$ to map the action of cooperation to 1 and the action of defection to 0. That is $\alpha(C) = 1$ and $\alpha(D) = 0$. The payoff of player i when the state of the game is \mathbf{a} is given by:

$$\pi_i(\mathbf{a}) = \sum_{j=1}^N \frac{\alpha(\mathbf{a}_j)b_j}{N} - \alpha(\mathbf{a}_i)c_i \quad (15)$$

The payoff difference that a player earns by unilaterally switching from C to D (or *vice-versa*) in the linear public goods game is independent of what the other co-players play in the game. That is, for every player i ,

$$\pi_i(D, \mathbf{q}) - \pi_i(C, \mathbf{q}) = c_i - \frac{b_i}{N} =: f_i(D, C) \quad (16)$$

is independent of co-players' actions \mathbf{q} . The linear public goods game is therefore an example of an additive game. This property of the game results in dominated strategies for players. Defection dominates cooperation when $c_i > b_i/N$ while cooperation dominates defection when $c_i < b_i/N$. Using proposition 2, one can derive the closed form expression for the stationary distribution of a N —player linear public goods game with two strategies.

Proposition 4. *When β is finite, the unique stationary distribution of the introspection dynamics for a N —player linear public goods game is given by:*

$$\mathbf{u}_{\mathbf{a}} = \prod_{j=1}^N \frac{1}{1 + e^{\text{sign}(\mathbf{a}_j)\beta f_j(D,C)}} \quad (17)$$

where,

$$\text{sign}(a) = \begin{cases} 1 & \text{if } a = C \\ -1 & \text{if } a = D \end{cases} \quad (18)$$

We use a simple example to demonstrate the above result. Consider a 3-player linear public goods game. All players provide a benefit of 2 units when they contribute to the public good ($b_1 = b_2 = b_3 = 2$). They differ however in their cost of cooperation. For player 1 and 2, the cost of cooperation is 1 unit

($c_1 = c_2 = 1$) while for the third player, the cost is slightly higher at 1.5 units ($c_3 = 1.5$). In the stationary distribution of the process with selection strength $\beta = 1$, the state where player 1 and 2 play C and player 3 plays D, \mathbf{u}_{CCD} , occurs with probability ≈ 0.112 (rounded up to three decimal places). We calculate this probability exactly with Eq. (17). The marginal probabilities that in the long run player 1 (or player 2) plays cooperation and player 3 plays defection are 0.4 and 0.7 respectively. That is, $\xi_{1:C} = \xi_{2:C} \approx 0.4$ and $\xi_{3:D} \approx 0.7$ (rounded upto a single decimal place). With the exact values (and also with the approximations here), one can confirm the factorizing property of the stationary distribution for additive games in this example (i.e., proposition 3). That is, $\mathbf{u}_{CCD} = \xi_{1:C} \cdot \xi_{2:C} \cdot \xi_{3:D}$.

First, we systematically the linear public goods game in which all players are symmetric. We consider a game with 4 players. All players have the same cost of cooperation ($c_i = c$) and they all provide the same benefit to the public goods ($b_i = b$). Since all players are identical, the states of the game can be enumerated by counting the number of cooperators in the state. There can only be 5 distinct states of the game (from 0 to 4 contributors). When the parameters of the game are such that defection dominates cooperation ($b = 2, c = 1$, Fig. ??a), the stationary distribution of the process at high β ($\beta = 5$), indicates an exponential decrease in the probability of a state with increase in the number of cooperators in the state. However, for intermediate and low β , the long-run outcomes of the introspection dynamics are qualitatively different to the outcomes at high β . Here, the state with 1 cooperator (or even 2 cooperators, depending on how small β is) is the most probable state in the long-run (Fig. ??b). Naturally, β plays an important role determining the overall cooperation frequency in the public goods game. When β is low, the average cooperation frequency varies weakly with the strength of the dilemma, $b/N - c$ (Fig. ??c). Even when the temptation to defect is high ($b/N - c \approx -2$), players cooperate with a non-zero probability in the long run (≈ 0.1). Similarly, when cooperation is highly beneficial ($b/N - c \approx 2$), players defect with some probability (≈ 0.2). At higher values of β , the stationary behaviour of the players is more responsive to the payoffs of the game and thus reflect an abrupt change around the parameter region where the game transitions from defection-dominating to cooperation-dominating ($b/N - c = 0$).

For studying the linear public goods game with asymmetric players, we consider a setup with 3 players. All the players differ in their cost of cooperation and the benefit they provide to the public goods. In this setup, the reference player's (player 2) cost and benefit values are 1 and 2 units respectively. Player 1 and player 3 differ from the reference player in opposite directions. For player 1, the cost and benefit are $1 + \delta_c$ and $2 + \delta_b$ respectively while for player 3, the cost and benefit are $1 - \delta_c$ and $2 - \delta_b$, respectively. First, we study a simple case in which players only vary in their cost of cooperation ($\delta_b = 0$ and $\delta_c = 0.5$, Fig ??a). In the long run, the relative cooperation frequency of players reflect their relative ability to cooperate. The player with the lowest cooperation cost (player 3), cooperates with the highest probability (and *vice-versa*). Similarly, when players only differ in their ability to produce the public good ($\delta_b = 1$ and $\delta_c = 0$, Fig ??b), their relative cooperation in the long run reflects the relative benefits they provide with their cooperation. The player with the highest benefit value (player 1) cooperates with the highest

probability. For this particular setup with three asymmetric players, defection dominates cooperation for player 1 if and only if $\delta_b < 1 + 3\delta_c$. For player 3, defection dominates cooperation only when $\delta_b > 3\delta_c - 1$. For the parameter values $b = 2, c = 1$, defection always dominates cooperation for the reference player. These regions in the $\delta_b - \delta_c$ parameter plane that correspond to defection dominating cooperation are circumscribed by white dashed lines in Fig. ??c. When players learn to play at high selection strength, β , their cooperation frequency in the long-run reflect the rational play (Fig. ??c). In the long run, the average cooperation frequency of the group is low if the asymmetry in the benefit value is bounded, $3\delta_c - 1 < \delta_b < 3\delta_c + 1$. This includes the case where players are symmetric ($\delta_b = \delta_c = 0$). A relatively high cooperation frequency is only assured if players are aligned in their asymmetries (i.e., either $\delta_b < 3\delta_c + 1$ or $\delta_b > 3\delta_c - 1$).

Games with two actions and their properties under introspection dynamics

In this section, we study the stationary properties of games with two actions. These games are not necessarily additive. First, we study the simplest case in which all players are symmetric. A N -player symmetric normal form game with two actions has the following properties:

1. $\mathbf{A}_1 = \mathbf{A}_2 = \dots = \mathbf{A}_N := \mathcal{A} := \{C, D\}$

2. For any $i, j \in \{1, 2, \dots, N\}$, $a \in \mathcal{A}$ and $\mathbf{b} \in \prod_{j=1}^N \mathcal{A}$:

$$\pi_i(a, \mathbf{b}) = \pi_j(a, \mathbf{b}) \quad (19)$$

Since players are symmetric, states can again be enumerated by counting the number of C players in the state. We denote the payoff of a C and D player in a state where there are j C players with $\pi_C(j)$ and $\pi_D(j)$ respectively. From the previous section, we adopt the mapping function $\alpha(\cdot)$ that maps the action C to 1 and the action D to 0. We use the notation $\mathcal{C}(\mathbf{a})$ to denote the number of cooperators in the state \mathbf{a} . That is,

$$\mathcal{C}(\mathbf{a}) := \sum_{j=1}^N \alpha(\mathbf{a}_j) \quad (20)$$

The stationary distribution of a two-action symmetric game under introspection dynamics can be explicitly computed using the following proposition,

Proposition 5. *When β is finite, the unique stationary distribution of the introspection dynamics for the N -player symmetric normal form game with two actions, $\mathcal{A} = \{C, D\}$, $(\mathbf{u}_{\mathbf{a}})_{\mathbf{a} \in \mathcal{A}^N}$, is given by:*

$$\mathbf{u}_{\mathbf{a}} = \frac{1}{\Gamma} \prod_{j=1}^{\mathcal{C}(\mathbf{a})} e^{-\beta f(j-1)} \quad (21)$$

where $f(j) := \pi^D(j) - \pi^C(j)$ is the payoff difference earned by switching the play from D and C when there are j co-players playing C. The term Γ is the normalization factor given by:

$$\Gamma = \sum_{\forall \mathbf{a}' \in \mathcal{A}^N} \prod_{j=1}^{C(\mathbf{a}')} e^{-\beta f(j-1)} \quad (22)$$

The number of unique states of the game can be reduced to $N + 1$ from 2^N due to symmetry. In the reduced state space, the state k corresponds to k players playing C and $N - k$ players playing D. Then, Proposition 5 can be reformulated by relabelling the states as follows,

Corollary 1. *When β is finite, the unique stationary distribution, $(\mathbf{u}_k)_{k \in \{0,1,\dots,N\}}$, of the introspection dynamics for the N -player symmetric normal form game with two actions, $\mathcal{A} = \{C, D\}$, is given by*

$$\mathbf{u}_k = \frac{1}{\Gamma} \cdot \binom{N}{k} \cdot \prod_{j=1}^k e^{-\beta f(j-1)} \quad (23)$$

where, k represents the number of C players in the state and $f(j) := \pi^D(j) - \pi^C(j)$ is the payoff difference earned by switching the play D and C when there are j co-players playing C. The term Γ is the normalization factor, given by,

$$\Gamma = \sum_{k=0}^N \binom{N}{k} \cdot \prod_{j=1}^k e^{-\beta f(j-1)} \quad (24)$$

The above lemma follows directly from Proposition 5. The key step is to count the number of states in the state space \mathcal{A}^N that corresponds to exactly k , C players (and therefore $N - k$, D players). This count is simply the binomial coefficient $\binom{N}{k}$. For details, see the proof in the Appendix. In this paper, we do not derive a closed form expression for the stationary distribution of introspection dynamics in asymmetric games with two actions. Instead, we compute it numerically using the explicit formula in Eq. (8). In the next section, we study the stationary behaviour of the introspection dynamics in the general public goods game (symmetric and asymmetric).

An example of a game with two actions: the general public goods game

To study general public goods game we adopt the framework of general social dilemmas from Hauert et al.³⁴. This framework introduces a parametric normal form game with symmetric players. The players have two actions: cooperation, C and defection, D. Here, we extend their framework to account for players with asymmetric payoffs. For player i , the cost of cooperation is c_i . The benefit that they can generate for the public good is b_i . However, in the general public goods game, the benefit of cooperation generated by a player is either synergized (or discounted) by a factor depending on the number of cooperators already in the group and the synergy/discount factor, w . If there are already $k - 1$ cooperators

in the group, the effective benefit provided to the public good by the k^{th} cooperating player (say i), is $w^{k-1}b_i$. In this setup, we assume that all the orders in which cooperators' benefits can be synergized (or discounted) are equally likely. We clarify with an example. Consider that there are three cooperators in the group: player i, j and k . The total benefit that they provide to the public good can be one of the six possibilities from $x + yw + zw^2$, where x, y and z are permutations of b_i, b_j and b_k . We assume that all the six possibilities are equally likely, and therefore, the expected benefit is given by $\bar{b}(1 + w + w^2)$ where $\bar{b} = (b_i + b_j + b_k)/3$.

The complete state space of the game is $\mathbf{A} = \{C, D\}^N$. The payoff of a defector in a state (D, \mathbf{a}_{-i}) and that of a cooperator in state (C, \mathbf{a}_{-i}) where $\mathbf{a}_{-i} \in \{C, D\}^{N-1}$ are respectively given by:

$$\pi_i^D(D, \mathbf{a}_{-i}) = \sum_{i=1}^N b_i \alpha(\mathbf{a}_i) \cdot \frac{1}{N} \cdot \left(1 + w + w^2 + \dots w^{\mathcal{C}(D, \mathbf{a}_{-i})-1}\right) \quad (25)$$

$$\pi_i^C(C, \mathbf{a}_{-i}) = \sum_{i=1}^N b_i \alpha(\mathbf{a}_i) \cdot \frac{1}{N} \cdot \left(1 + w + w^2 + \dots w^{\mathcal{C}(C, \mathbf{a}_{-i})-1}\right) - c_i \quad (26)$$

where $\mathcal{C}(a, \mathbf{a}_{-i})$ is the number of cooperators in the state (a, \mathbf{a}_{-i}) and $\alpha(\cdot)$ is the mapping function that maps the actions C and D to 1 and 0 respectively. Note that the number of cooperators in the two states are related as: $\mathcal{C}(D, \mathbf{a}_{-i}) = \mathcal{C}(C, \mathbf{a}_{-i}) - 1$. In the symmetric case, where every player has the same cost and benefit values, the payoffs of a defector and a cooperator in a group with k cooperators can be simply written as:

$$\pi_i^D(k) = \frac{b}{N}(1 + w + w^2 + \dots + w^{k-1}) \quad (27)$$

$$\pi_i^C(k) = \pi_i^D(k) - c \quad (28)$$

The parameter w represents the nature of the public goods. The nature is linear when $w = 1$. Every additional cooperator's contribution is as valuable as the benefit that they can generate. When $w < 1$, the effective value of every new cooperator goes down by a factor of w (compared to the last cooperator). The public goods is said to be discounting in that case. On the other hand when $w > 1$, every new contributor is more valuable than the previous one. These public goods are called synergistic. For the symmetric case, the relationship between the cost to benefit ration cN/b and the discount/synergy factor determines the type of social dilemma arising from this general public goods game. In principle, this framework can produce generalizations of the prisoner's dilemma (defection dominating cooperation), the snowdrift game (no dominance but coexistence under the replicator dynamics), the stag-hunt game

(no dominance but existence of an internal equilibrium that is unstable under the replicator dynamics) and the harmony game (cooperation dominating defection) with respect to its evolutionary trajectories. For more details see Hauert et al.³⁴. We study the long term stationary behaviour of players when they learn through introspection dynamics. First we focus on only the symmetric version of this game.

To compute the stationary distribution of the symmetric general public goods game (PG), we use the closed form expression in Eq. (23). We consider that every player in a 4-player game can generate a benefit b of value 2. We study an introspection process with selection strength $\beta = 5$. Before exploring the $c - w - N$ parameter space, we study four specific cases. In two of these cases, the public good is discounted ($w = 0.5$) and in two other cases, the public good is synergistic ($w = 1.5$). For each case we consider two sub-cases: one, in which cost is high ($c = 1$) and two, when cost is low ($c = 0.2$). The four parameter combinations are chosen such that each of them represents a unique social dilemma under the replicator dynamics. When the PG is discounted ($w = 0.5$) and costs are high ($c = 1$) defection dominates cooperation (the prisoner's dilemma). When the PG is discounted ($w = 0.5$) but costs are low ($c = 0.2$), the coexistence behaviour similar to the snowdrift game appears from the replicator equation. When the PG is synergistic ($w = 1.5$) and costs are high ($c = 1$), both cooperation and defection are locally stable equilibria. In addition there is also an internal equilibrium which is unstable. This represents the stag-hunt game. And finally when both the PG is synergistic and costs are low, cooperation dominates defection. This situation represents the harmony game. The stationary behaviour of the players from the introspection dynamics in these games reflect the outcomes from the replicator dynamics (Fig. ??a). Since selection strength is chosen to be intermediate, in some of these cases, players play actions that are not optimal for the dilemma with positive probability in the long run. For example, even when the parameters of the game make cooperation to be the dominated strategy, there is a single cooperator in the group with a sizeable chance ($\approx 20\%$) in the long run. When the parameters of the game reflect the stag-hunt dilemma, players are more likely to coordinate their actions in the long run. The probabilities that the whole group plays C or D is higher than the probability that there is a group with mixed play. In contrast, when the parameters reflect the snowdrift game, we get the opposite effect. In the long run, mixed groups are more likely than homogeneous groups. In fact, the group with exactly 2 cooperators and 2 defectors is the most likely outcome in the long run. Finally, when the parameters of the game make defection the dominated action, all players in the group learn to play cooperation in the long run.

The average cooperation frequency of the group in the long run are shown in the $c - w$ and $N - w$ parameter planes in Fig ??b. First, let us consider the case when the group size is fixed at 4 players (the $c - w$ plane in Fig ??b). In that case if the cost of cooperation is too high, the average cooperation rate is negligible and does not change with the change in the nature of the public good. When the cost is not restrictively high, the discount/synergy parameter, w , determines the frequency with which players cooperate in the long run. A high w for the public good would result in high cooperation in the long run (and *vice-versa*). Next we consider the case where the cost of cooperation is fixed (the $N - w$ plane in Fig.

??b). It is fixed to a value such that there is full cooperation in the long run in a group of arbitrary size when $w > 1$. In this case, when the public good is discounted, group size N and the discounting factor w jointly determine the cooperation frequency in the long run. For a fixed value of $w(< 1)$, cooperation rates fall with larger group sizes.

Discussions and conclusion

Introspection dynamics is a rather broad model. Here, we mainly focused on general cases to the detriment of more realistic, complex examples. Nevertheless, in the last section, we give one step forward in that direction, by studying a 2-stage game, where individuals are allowed to reward their co-players condition on previous actions. Clearly, there are a number of ways in which we can further employ our model. For example, other researchers recently studied multiplayer games considering multiple games played concurrently⁸, fluctuating environments⁹, continuous strategies⁷, or repeated interactions⁷.

Finally, for the sake of mathematical tractability, introspection dynamics, as we present it here, does not consider a population, not to mention complex population structures⁴⁵. However, it can be equally applicable to population models. In that case, players obtain payoffs either from well-mixed or network-bounded interactions, as usual, but update their strategies introspectively.

Appendix: Proofs

Proof. Proof of Proposition 2

Since β is finite, the stationary distribution $\mathbf{u} = (\mathbf{u}_{\mathbf{a}})_{\mathbf{a} \in \mathbf{A}}$ of the process is unique by Proposition 1. The stationary distribution also satisfies the equalities in Eq. (6) and (7). Before continuing through the remainder of the proof, we introduce some short-cut notation that we will be using:

$$I_{\mathbf{b}} := I(\mathbf{b}, \mathbf{a}), \quad \text{iff} \quad \mathbf{b} \in \text{Neb}(\mathbf{a}) \quad (29)$$

$$\tau_{j:\mathbf{a}_j} := \frac{1}{\sum_{\mathbf{a}' \in \mathbf{A}_j} e^{\beta f_j(\mathbf{a}', \mathbf{a}_j)}} \quad (30)$$

In order to show that the candidate stationary distribution, as proposed in Eq. (12) is the stationary distribution of the process, we need to show that the following are true:

$$\mathbf{T}_{a,a}\mathbf{u}_a + \sum_{b \neq a} \mathbf{T}_{b,a}\mathbf{u}_b = \mathbf{u}_a \quad \forall a \in \mathbf{A} \quad (31)$$

$$\sum_{a \in \mathbf{A}} \mathbf{u}_a = 1 \quad (32)$$

Using our short-cut notation τ and the expression for our candidate stationary distribution in Eq. (12), we can express the stationary distribution as:

$$\mathbf{u}_a = \prod_{j=1}^N \tau_{j:a_j} \quad (33)$$

Using this expression, the left hand side of Eq. (31) can be simplified further with the steps:

$$\mathbf{T}_{a,a}\mathbf{u}_a + \sum_{b \neq a} \mathbf{T}_{b,a}\mathbf{u}_b \quad (34)$$

$$= \left(1 - \frac{1}{N} \sum_{b \in \text{Neb}(a)} \frac{1}{m_{I_b} - 1} \cdot p_{a_{I_b} \rightarrow b_{I_b}} \cdot \mathbf{u}_a \right) + \sum_{b \in \text{Neb}(a)} \frac{1}{m_{I_b} - 1} \cdot p_{b_{I_b} \rightarrow a_{I_b}} \cdot \mathbf{u}_b \quad (35)$$

$$= \mathbf{u}_a + \frac{1}{N} \sum_{b \in \text{Neb}(a)} \left(\prod_{k \neq I_b} \tau_{k:a_k} \right) \left(p_{b_{I_b} \rightarrow a_{I_b}} \cdot \tau_{I_b:a_{I_b}} - p_{a_{I_b} \rightarrow b_{I_b}} \cdot \tau_{I_b:b_{I_b}} \right) \cdot \left(\frac{1}{m_{I_b} - 1} \right) \quad (36)$$

For an additive game, the expressions for $p_{b_{I_b} \rightarrow a_{I_b}}$ and $p_{a_{I_b} \rightarrow b_{I_b}}$ can be simply written as:

$$p_{b_{I_b} \rightarrow a_{I_b}} = \frac{1}{1 + e^{\beta f_{I_b}(b_{I_b}, a_{I_b})}} \quad (37)$$

$$p_{a_{I_b} \rightarrow b_{I_b}} = \frac{1}{1 + e^{\beta f_{I_b}(a_{I_b}, b_{I_b})}} \quad (38)$$

Using the above expressions and the expression for τ in Eq. (30), it can be shown that:

$$\left(p_{b_{I_b} \rightarrow a_{I_b}} \cdot \tau_{I_b:a_{I_b}} - p_{a_{I_b} \rightarrow b_{I_b}} \cdot \tau_{I_b:b_{I_b}} \right) = 0 \quad (39)$$

After plugging the equality in Eq. (39) into Eq. (36), we see that the left hand side of Eq. (31) simplifies to \mathbf{u}_a . Now, to complete the proof we must check if Eq. (32) holds for our candidate distribution. Sum-

ming up the elements of the stationary distribution \mathbf{u}_a for all states $a \in \mathbf{A}$:

$$\sum_{a \in \mathbf{A}} \mathbf{u}_a = \sum_{a \in \mathbf{A}} \prod_{k=1}^N \tau_{k:a_k} = \sum_{a \in \mathbf{A}} \frac{\prod_{k=1}^N e^{\beta \pi_k(a_k, \mathbf{q}_{-k})}}{\prod_{k=1}^N \sum_{a' \in \mathbf{A}_k} e^{\beta \pi_k(a', \mathbf{q}_{-k})}} \quad (40)$$

where $\mathbf{q}_{-1}, \mathbf{q}_{-2}, \dots, \mathbf{q}_{-N}$ are fixed tuples from $\mathbf{A}_{-1}, \mathbf{A}_{-2}, \dots, \mathbf{A}_{-N}$ respectively. The denominator in the above expression can be taken out completely from the first sum. That is,

$$\sum_{a \in \mathbf{A}} \mathbf{u}_a = \sum_{a \in \mathbf{A}} \frac{\prod_{k=1}^N e^{\beta \pi_k(a_k, \mathbf{q}_{-k})}}{\prod_{k=1}^N \sum_{a' \in \mathbf{A}_k} e^{\beta \pi_k(a', \mathbf{q}_{-k})}} \quad (41)$$

$$= \left(\prod_{k=1}^N \left(e^{\beta \pi_k(a_{k,1}, \mathbf{q}_{-k})} + \dots + e^{\beta \pi_k(a_{k,m_k}, \mathbf{q}_{-k})} \right) \right)^{-1} \cdot \left(\sum_{a \in \mathbf{A}} \prod_{k=1}^N e^{\beta \pi_k(a_k, \mathbf{q}_{-k})} \right) \quad (42)$$

$$(43)$$

Multiplying out the sums in the denominator of the above expression, we get that:

$$\sum_{a \in \mathbf{A}} \mathbf{u}_a = \left(\prod_{k=1}^N \left(e^{\beta \pi_k(a_{k,1}, \mathbf{q}_{-k})} + \dots + e^{\beta \pi_k(a_{k,m_k}, \mathbf{q}_{-k})} \right) \right)^{-1} \cdot \left(\sum_{a \in \mathbf{A}} \prod_{k=1}^N e^{\beta \pi_k(a_k, \mathbf{q}_{-k})} \right) \quad (44)$$

$$= \left(\sum_{a \in \mathbf{A}} \prod_{k=1}^N e^{\beta \pi_k(a_k, \mathbf{q}_{-k})} \right)^{-1} \left(\sum_{a \in \mathbf{A}} \prod_{k=1}^N e^{\beta \pi_k(a_k, \mathbf{q}_{-k})} \right) = 1 \quad (45)$$

Therefore, the stationary distribution sums up to 1. The candidate distribution we propose for the additive game is the unique stationary distribution of the process. \square

Proof. **Proof of Proposition 3**

Just like the previous proof, we will denote $\mathbf{p}_{-1}, \mathbf{p}_{-2}, \dots, \mathbf{p}_{-N}$ as fixed tuples from $\mathbf{A}_{-1}, \mathbf{A}_{-2}, \dots, \mathbf{A}_{-N}$ respectively. In the steps below, we always decompose the expression $f_j(a, b)$ to $\pi_j(a, \mathbf{p}_{-j}) - \pi_j(b, \mathbf{p}_{-j})$.

When $\mathbf{u} = (\mathbf{u}_{\mathbf{a}})_{\mathbf{a} \in \mathbf{A}}$ is the unique stationary distribution of the N –player additive game under finite selection introspection dynamics, it is given by the closed form expression in Eq. (12). We use this expression to calculate the marginal distribution of actions played at a particular state \mathbf{a} , $(\xi_{j:\mathbf{a}_j})_{j \in \{1,2,\dots,N\}}$.

$$\xi_{j:\mathbf{a}_j} = \sum_{\mathbf{q} \in \mathbf{A}_{-j}} \mathbf{u}_{(\mathbf{a}_j, \mathbf{q})} \quad (46)$$

$$= \sum_{\mathbf{q} \in \mathbf{A}_{-j}} \left(\sum_{a' \in \mathbf{A}_j} e^{\beta f_j(a', \mathbf{a}_j)} \right)^{-1} \prod_{k \neq j} \left(\sum_{a' \in \mathbf{A}_k} e^{\beta f_k(a', \mathbf{q}_k)} \right)^{-1} \quad (47)$$

$$= \left(\prod_{k=1}^N \sum_{a' \in \mathbf{A}_k} e^{\beta \pi_k(a', \mathbf{p}_{-k})} \right)^{-1} \cdot e^{\beta \pi_j(\mathbf{a}_j, \mathbf{p}_{-j})} \cdot \left(\sum_{\mathbf{q} \in \mathbf{A}_{-j}} \prod_{k \neq j} e^{\beta \pi_k(\mathbf{q}_k, \mathbf{l}_{-k})} \right) \quad (48)$$

$$= \left(\sum_{a' \in \mathbf{A}_j} e^{\beta \pi_j(a', \mathbf{p}_{-j})} \right)^{-1} \cdot e^{\beta \pi_j(\mathbf{a}_j, \mathbf{p}_{-j})} \cdot \left(\prod_{k \neq j} \sum_{a' \in \mathbf{A}_k} e^{\beta \pi_k(a', \mathbf{p}_{-k})} \right)^{-1} \cdot \left(\sum_{\mathbf{q} \in \mathbf{A}_{-j}} \prod_{k \neq j} e^{\beta \pi_k(\mathbf{q}_k, \mathbf{p}_{-k})} \right) \quad (49)$$

$$= \left(\sum_{a' \in \mathbf{A}_j} e^{\beta \pi_j(a', \mathbf{p}_{-j})} \right)^{-1} \cdot e^{\beta \pi_j(\mathbf{a}_j, \mathbf{p}_{-j})} \cdot \left(\sum_{\mathbf{q} \in \mathbf{A}_{-j}} \prod_{k \neq j} e^{\beta \pi_k(\mathbf{q}_k, \mathbf{p}_{-k})} \right)^{-1} \cdot \left(\sum_{\mathbf{q} \in \mathbf{A}_{-j}} \prod_{k \neq j} e^{\beta \pi_k(\mathbf{q}_k, \mathbf{p}_{-k})} \right) \quad (50)$$

$$= \left(\sum_{a' \in \mathbf{A}_j} e^{\beta(\pi_j(a', \mathbf{p}_{-j}) - \pi_j(\mathbf{a}_j, \mathbf{p}_{-j}))} \right)^{-1} \quad (51)$$

$$= \sum_{a' \in \mathbf{A}_j} e^{\beta f_j(a', \mathbf{a}_j)} \quad (52)$$

The interchange of the sum and the product between the expressions in Eq. (49) and Eq. (50) can be carried out by observing that when all the sums are multiplied out, one is left with sums of terms, each of which is a exponential with power equal to sum of payoffs that co-players of j (here k) receive when they play their respective strategies from \mathbf{q} (that is \mathbf{q}_k) against co-players that play \mathbf{p}_{-k} . Using the expression in Eq. (52), we can confirm that for additive games, the product of the marginals is the stationary distribution,

$$\prod_{j=1}^N \xi_{j:\mathbf{a}_j} = \mathbf{u}_{\mathbf{a}} \quad (53)$$

□

Proof. **Proof of Proposition 4**

Since we have demonstrated that the linear public goods game is an additive game, the proof of this theorem can be performed by directly using proposition 2. Here, we provide an independent proof. The idea behind this proof is identical to the proof of proof2.

Using proposition 1, we can confirm that since β is finite, the process will have a unique stationary distribution. Before continuing with the rest of the proof where we show that our candidate stationary distribution is *the* unique stationary distribution, we define the following short-cut notations for the ease of the proof:

$$\bar{\mathbf{a}}_j := \{D, C\} \setminus \{\mathbf{a}_j\} \quad (54)$$

$$p_j := \frac{1}{1 + e^{\beta f_j(D, C)}} \quad (55)$$

In addition we introduce a mapping function $\alpha(\cdot)$ which maps the action C to 1 and the action D to 0. That is $\alpha(C) := 1$ and $\alpha(D) := 0$. Using these notations and Eq. (1) and (16) and utilizing our shortcut notation from above, we can write the probability that a player j updates to \mathbf{a}_j from $\bar{\mathbf{a}}_j$ while their co-players play \mathbf{a}_{-j} as:

$$p_{\bar{\mathbf{a}}_j \rightarrow \mathbf{a}_j}(\mathbf{a}_{-j}) = p_j \text{sign}(\mathbf{a}_j) + \alpha(\bar{\mathbf{a}}_j) \quad (56)$$

The candidate stationary distribution \mathbf{u} given in Eq. (17) can be written down using our short-cut notation as:

$$\mathbf{u}_{\mathbf{a}} = \prod_{k=1}^N p_k \text{sign}(\mathbf{a}_k) + \alpha(\bar{\mathbf{a}}_k) \quad (57)$$

This stationary distribution must satisfy the following properties, which are also given in Eq (6) and (7):

$$\mathbf{u}_{\mathbf{a}} = \mathbf{T}_{\mathbf{a}, \mathbf{a}} \mathbf{u}_{\mathbf{a}} + \sum_{\mathbf{b} \neq \mathbf{a}} \mathbf{T}_{\mathbf{b}, \mathbf{a}} \mathbf{u}_{\mathbf{b}} \quad (58)$$

$$\sum_{\mathbf{a} \in \mathbf{A}} \mathbf{u}_{\mathbf{a}} = 1 \quad (59)$$

Where, the terms in the right hand side of Eq. 58 can be simplified using Eq. (1) and (5) as follows:

$$\mathbf{T}_{\mathbf{a},\mathbf{a}} = 1 - \sum_{k=1}^N \mathbf{T}_{(\mathbf{a}_k, \mathbf{a}_{-k}), (\bar{\mathbf{a}}_k, \mathbf{a}_{-k})} = 1 - \frac{1}{N} \sum_{k=1}^N p_k \text{sign}(\bar{\mathbf{a}}_k) + \alpha(\mathbf{a}_k) \quad (60)$$

and additionally, also using Eq. (57) the second term can be simplified too:

$$\sum_{\mathbf{b} \neq \mathbf{a}} \mathbf{T}_{\mathbf{b},\mathbf{a}} \mathbf{u}_{\mathbf{b}} = \sum_{k=1}^N \mathbf{T}_{(\bar{\mathbf{a}}_k, \mathbf{a}_{-k}), (\mathbf{a}_k, \mathbf{a}_{-k})} \mathbf{u}_{(\bar{\mathbf{a}}_k, \mathbf{a}_{-k})} \quad (61)$$

$$= \frac{1}{N} \sum_{k=1}^N (p_k \text{sign}(\mathbf{a}_k) + \alpha(\bar{\mathbf{a}}_k)) \mathbf{u}_{(\bar{\mathbf{a}}_k, \mathbf{a}_{-k})} \quad (62)$$

$$= \frac{\mathbf{u}_{\mathbf{a}}}{N} \sum_{k=1}^N p_k \text{sign}(\bar{\mathbf{a}}_k) + \alpha(\mathbf{a}_k) \quad (63)$$

Now, using Eq. (60) and (63) one can show that the right hand side of Eq. (58) is the element of the stationary distribution, corresponding to the state \mathbf{a} , $\mathbf{u}_{\mathbf{a}}$. Now, to complete the proof, we must show that Eq. (59) is also true for our candidate stationary distribution. This can be done by decomposing the sum of the elements of the stationary distribution as follows:

$$\sum_{\mathbf{a} \in \mathbf{A}} u_{\mathbf{a}} = \sum_{\mathbf{a} \in \mathbf{A}} \prod_{k=1}^N p_k \text{sign}(\mathbf{a}_k) + \alpha(\bar{\mathbf{a}}_k) \quad (64)$$

$$= \sum_{\mathbf{a} \in \mathbf{A}_{-N}} (1 - p_N) \prod_{k=1}^{N-1} p_k \text{sign}(\mathbf{a}_k) + \alpha(\bar{\mathbf{a}}_k) + p_N \prod_{k=1}^{N-1} p_k \text{sign}(\mathbf{a}_k) + \alpha(\bar{\mathbf{a}}_k) \quad (65)$$

$$= \sum_{\mathbf{a} \in \mathbf{A}_{-N}} \prod_{k=1}^{N-1} p_k \text{sign}(\mathbf{a}_k) + \alpha(\bar{\mathbf{a}}_k) \quad (66)$$

When the above decomposition is performed $N - 1$ more times, the sum of the right hand side becomes 1. This proves that the candidate stationary distribution is also a probability distribution. \square

Proof. Proof of Proposition 5

By construction, the candidate stationary distribution given by Eq. (21) and Eq. (22) is a probability distribution since it satisfies the condition in Eq. (7) and for any state \mathbf{a} , $\mathbf{u}_{\mathbf{a}}$ is between 0 and 1. Moreover, since β is finite, by proposition 1, the process will have a unique stationary distribution. To show that the candidate stationary distribution is the unique stationary distribution, we need to check if for this process, $\mathbf{uT} = \mathbf{u}$. That is, the condition in Eq. (58) must hold for all states \mathbf{a} . We re-introduce some

notations that we will use in this proof:

$$\bar{\mathbf{a}}^j := \{D, C\} \setminus \{\mathbf{a}_j\} \quad (67)$$

$$\alpha(a) := \begin{cases} 1 & \text{if } a = C \\ 0 & \text{if } a = D \end{cases} \quad (68)$$

$$\mathcal{C}(\mathbf{a}) = \sum_{j=1}^N \alpha(\mathbf{a}_j) \quad (69)$$

For this process, the first term in the right hand side of Eq. (58) can be simplified as:

$$\mathbf{u}_{\mathbf{a}} \mathbf{T}_{\mathbf{a}, \mathbf{a}} = \mathbf{u}_{\mathbf{a}} - \mathbf{u}_{\mathbf{a}} \sum_{k=1}^N \mathbf{T}_{(\mathbf{a}_k, \mathbf{a}_{-k}), (\bar{\mathbf{a}}_k, \mathbf{a}_{-k})} \quad (70)$$

$$= \mathbf{u}_{\mathbf{a}} - \frac{\mathbf{u}_{\mathbf{a}}}{N} \sum_{k=1}^N \frac{1}{1 + e^{\text{sign}(\bar{\mathbf{a}}_k) \beta f(N_k)}} \quad (71)$$

Where, the function $\text{sign}(\cdot)$ is defined as in Eq. (18) and $f(j)$ is the difference in payoffs between playing D and C when there are j co-players playing C. The term N_k represents the number of co-players of k that play C in state \mathbf{a} . That is,

$$N_k := \sum_{j \neq k} \alpha(\mathbf{a}_j) \quad (72)$$

The second term in the right hand side of Eq. 58 can be simplified as,

$$\sum_{\mathbf{b} \neq \mathbf{a}} \mathbf{T}_{\mathbf{b}, \mathbf{a}} \mathbf{u}_{\mathbf{b}} = \sum_{k=1}^N \mathbf{T}_{(\bar{\mathbf{a}}_k, \mathbf{a}_{-k}), (\mathbf{a}_k, \mathbf{a}_{-k})} \mathbf{u}_{(\bar{\mathbf{a}}_k, \mathbf{a}_{-k})} \quad (73)$$

$$= \frac{1}{N\Gamma} \sum_{k=1}^N \mathbf{T}_{(\bar{\mathbf{a}}_k, \mathbf{a}_{-k}), (\mathbf{a}_k, \mathbf{a}_{-k})} \prod_{j=1}^{\mathcal{C}((\bar{\mathbf{a}}_k, \mathbf{a}_{-k}))} e^{-\beta f(j-1)} \quad (74)$$

$$= \frac{1}{N\Gamma} \sum_{k=1}^N \mathbf{T}_{(\bar{\mathbf{a}}_k, \mathbf{a}_{-k}), (\mathbf{a}_k, \mathbf{a}_{-k})} \left(\prod_{j=1}^{N_k} e^{-\beta f(j-1)} \right) \cdot e^{-\beta \alpha(\bar{\mathbf{a}}_k) f(-\alpha(\mathbf{a}_k) + N_k)} \quad (75)$$

Between the steps, Eq. 74 and 75, we took out one term from the product that is present in our candidate distribution. This term accounts for the k^{th} players action in the neighbouring state $(\bar{\mathbf{a}}_k, \mathbf{a}_{-k})$ of \mathbf{a} . For simplicity, we replace $\mathbf{T}_{(\bar{\mathbf{a}}_k, \mathbf{a}_{-k}), (\mathbf{a}_k, \mathbf{a}_{-k})}$ with just \mathbf{T} in the next steps. We continue the simplification of

Eq. (75) in the later steps by introducing terms that cancel each other.

$$\sum_{\mathbf{b} \neq \mathbf{a}} \mathbf{T}_{\mathbf{b}, \mathbf{a}} \mathbf{u}_{\mathbf{b}} = \frac{1}{N\Gamma} \sum_{k=1}^N \mathbf{T} \cdot \left(\prod_{j=1}^{N_k} e^{-\beta f(j-1)} \right) \cdot \frac{e^{-\beta \alpha(\bar{\mathbf{a}}_k) f(-\alpha(\mathbf{a}_k) + N_k)}}{e^{-\beta \alpha(\mathbf{a}_k) f(-\alpha(\bar{\mathbf{a}}_k) + N_k)}} \cdot e^{-\beta \alpha(\mathbf{a}_k) f(-\alpha(\bar{\mathbf{a}}_k) + N_k)} \quad (76)$$

The newly introduced term in Eq. (76) can be taken inside the product. Note that this term is 0 if the k^{th} player plays D in the state \mathbf{a} . When this term is taken inside the product bracket, products of exponent $e^{-\beta f(j-1)}$ can be performed for j ranging from 1 to the number of cooperators in state \mathbf{a} , $\mathcal{C}(\mathbf{a})$. This product is then the candidate stationary distribution probability $\mathbf{u}_{\mathbf{a}}$. That is,

$$\sum_{\mathbf{b} \neq \mathbf{a}} \mathbf{T}_{\mathbf{b}, \mathbf{a}} \mathbf{u}_{\mathbf{b}} = \frac{1}{N\Gamma} \sum_{k=1}^N \mathbf{T} \cdot \left(\prod_{j=1}^{N_k} e^{-\beta f(j-1)} \cdot e^{-\beta \alpha(\mathbf{a}_k) f(-\alpha(\bar{\mathbf{a}}_k) + N_k)} \right) \cdot \frac{e^{-\beta \alpha(\bar{\mathbf{a}}_k) f(-\alpha(\mathbf{a}_k) + N_k)}}{e^{-\beta \alpha(\mathbf{a}_k) f(-\alpha(\bar{\mathbf{a}}_k) + N_k)}} \quad (77)$$

$$= \frac{1}{N} \sum_{k=1}^N \mathbf{T} \cdot \left(\frac{1}{\Gamma} \prod_{j=1}^{\mathcal{C}(\mathbf{a})} e^{-\beta f(j-1)} \right) \cdot \frac{e^{-\beta \alpha(\bar{\mathbf{a}}_k) f(-\alpha(\mathbf{a}_k) + N_k)}}{e^{-\beta \alpha(\mathbf{a}_k) f(-\alpha(\bar{\mathbf{a}}_k) + N_k)}} \quad (78)$$

$$= \frac{1}{N} \sum_{k=1}^N \mathbf{T}_{(\bar{\mathbf{a}}_k, \mathbf{a}_{-k}), (\mathbf{a}_k, \mathbf{a}_{-k})} \cdot \mathbf{u}_{\mathbf{a}} \cdot \frac{e^{-\beta \alpha(\bar{\mathbf{a}}_k) f(-\alpha(\mathbf{a}_k) + N_k)}}{e^{-\beta \alpha(\mathbf{a}_k) f(-\alpha(\bar{\mathbf{a}}_k) + N_k)}} \quad (79)$$

The fraction inside the sum in Eq. (79) can be simplified as follows leading to further simplification of Eq. (79):

$$\sum_{\mathbf{b} \neq \mathbf{a}} \mathbf{T}_{\mathbf{b}, \mathbf{a}} \mathbf{u}_{\mathbf{b}} = \frac{1}{N} \sum_{k=1}^N \mathbf{T}_{(\bar{\mathbf{a}}_k, \mathbf{a}_{-k}), (\mathbf{a}_k, \mathbf{a}_{-k})} \cdot \mathbf{u}_{\mathbf{a}} \cdot e^{\text{sign}(\mathbf{a}_k) \beta f(N_k)} \quad (80)$$

In Eq. (80) we can replace the element of the transition matrix $\mathbf{T}_{(\bar{\mathbf{a}}_k, \mathbf{a}_{-k}), (\mathbf{a}_k, \mathbf{a}_{-k})}$ by using the following:

$$\mathbf{T}_{(\bar{\mathbf{a}}_k, \mathbf{a}_{-k}), (\mathbf{a}_k, \mathbf{a}_{-k})} = \frac{1}{1 + e^{\text{sign}(\mathbf{a}_k) \beta f(N_k)}} \quad (81)$$

Using the expression for the transition matrix element from Eq. (81) into Eq. (80) and by using Eq. (71),

we can simplify further:

$$\sum_{\mathbf{b} \neq \mathbf{a}} \mathbf{T}_{\mathbf{b}, \mathbf{a}} \mathbf{u}_{\mathbf{b}} = \frac{\mathbf{u}_{\mathbf{a}}}{N} \sum_{k=1}^N \frac{1}{1 + e^{\text{sign}(\mathbf{a}_k) \beta f(N_k)}} \cdot e^{\text{sign}(\mathbf{a}_k) \beta f(N_k)} \quad (82)$$

$$= \frac{\mathbf{u}_{\mathbf{a}}}{N} \sum_{k=1}^N \frac{1}{1 + e^{\text{sign}(\bar{\mathbf{a}}_k) \beta f(N_k)}} \quad (83)$$

$$= \mathbf{u}_{\mathbf{a}} - \mathbf{u}_{\mathbf{a}} \mathbf{T}_{\mathbf{a}, \mathbf{a}} \quad (84)$$

The final step in the previous simplification shows that Eq. (58) holds for any $\mathbf{a} \in \{C, D\}^N$. Therefore, the candidate distribution we propose in Eq. (21) is the unique stationary distribution of the symmetric N -player game with two strategies. \square

Proof. Proof of Corollary 1

To show this result we count how many states are identical to a state $\mathbf{a} \in \{C, D\}^N$ in a symmetric game. When players are symmetric in a two-strategy game, states can be enumerated by counting the number of C players in that state. This can also be confirmed by the expression of the stationary distribution in Eq. 21. Two distinct states \mathbf{a}, \mathbf{a}' having the same number of cooperators (i.e., $\mathcal{C}(\mathbf{a}') = \mathcal{C}(\mathbf{a})$), have the same stationary distribution probability (i.e., $\mathbf{u}_{\mathbf{a}'} = \mathbf{u}_{\mathbf{a}}$).

In a game with N players, there can be k players playing C in exactly $\binom{N}{k}$ ways. As argued before, all of these states are identical and are also equiprobable in the stationary distribution. Therefore, the stationary distribution probability of having k , C players, \mathbf{u}_k , is,

$$\mathbf{u}_k = \sum_{\forall \mathbf{a}, \mathcal{C}(\mathbf{a})=k} \mathbf{u}_{\mathbf{a}} = \frac{1}{\Gamma} \binom{N}{k} \prod_{j=1}^k e^{-\beta f(j-1)} \quad (85)$$

Where the normalization factor Γ can also be simplified as:

$$\Gamma = \sum_{k=0}^N \binom{N}{k} \prod_{j=1}^k e^{-\beta f(j-1)} \quad (86)$$

\square

Supplementary References

References

- [1] Günther Palm. Evolutionary stable strategies and game dynamics for n-person games. *Journal of Mathematical Biology*, 19(3):329–334, Jul 1984.
- [2] B. Skyrms. *The Stag-Hunt Game and the Evolution of Social Structure*. Cambridge University Press, Cambridge, 2003.
- [3] J. M. Pacheco, F. C. Santos, M. O. Souza, and B. Skyrms. Evolutionary dynamics of collective action in n-person stag hunt dilemmas. *Proceedings of the Royal Society B*, 276:315–321, 2009.
- [4] Marco Archetti, István Scheuring, Moshe Hoffman, Megan E Frederickson, Naomi E Pierce, and Douglas W Yu. Economic game theory for mutualism and cooperation. *Ecology Letters*, 14(12):1300–1312, 2011.
- [5] Marco Archetti and István Scheuring. Review: Game theory of public goods in one-shot social dilemmas without assortment. *Journal of Theoretical Biology*, 299(0):9–20, 2012.
- [6] C. S. Gokhale and A. Traulsen. Evolutionary multiplayer games. *Dynamic Games and Applications*, 4:468–488, 2014.
- [7] C. Hilbe, B. Wu, A. Traulsen, and M. A. Nowak. Evolutionary performance of zero-determinant strategies in multiplayer games. *Journal of Theoretical Biology*, 374:115–124, 2015.
- [8] V. R. Venkateswaran and C. S. Gokhale. Evolutionary dynamics of complex multiple games. *Proceedings of the Royal Society B*, 286:20190900, 2019.
- [9] Saptarshi Pal and Christian Hilbe. Reputation effects drive the joint evolution of cooperation and social rewarding. *Nature Communications*, 13(1):5928, 2022.
- [10] M. C. Couto, J. M. Pacheco, and F. C. Santos. Governance of risky public goods under graduated punishment. *Journal of Theoretical Biology*, 505:110423, 2020.
- [11] C. S. Gokhale and A. Traulsen. Strategy abundance in evolutionary many-player games with multiple strategies. *Journal of Theoretical Biology*, 238:180–191, 2011.
- [12] J. Peña. Group size diversity in public goods games. *Evolution*, 66:623–636, 2012.
- [13] M. Broom, K. Pattni, and J. Rychtář. Generalized social dilemmas: The evolution of cooperation in populations with variable group size. *Bulletin of Mathematical Biology*, 81:4643–4674, 2019.
- [14] Peter D. Taylor. Evolutionarily stable strategies with two types of player. *Journal of Applied Probability*, 16(1):76–83, 1979.

- [15] P. Schuster and K. Sigmund. Coyness, philandering and stable strategies. *Animal Behaviour*, 29:186–192, 1981.
- [16] A. Gaunersdorfer, J. Hofbauer, and K. Sigmund. The dynamics of asymmetric games. *Theoretical Population Biology*, 29:345–357, 1991.
- [17] J. Hofbauer. Evolutionary dynamics for bimatrix games: A Hamiltonian system? *Journal of Mathematical Biology*, 34:675–688, 1996.
- [18] Josef Hofbauer and Ed Hopkins. Learning in perturbed asymmetric games. *Games and Economic Behavior*, 52(1):133–152, 2005.
- [19] H. Ohtsuki. Stochastic evolutionary dynamics of bimatrix games. *Journal of Theoretical Biology*, 264:136–142, 2010.
- [20] Alex McAvoy and Christoph Hauert. Asymmetric Evolutionary Games. *PLoS Computational Biology*, 11(8):1–26, 2015.
- [21] C. Veller and L. K. Hayward. Finite-population evolution with rare mutations in asymmetric games. *Journal of Economic Theory*, 162:93–113, 2016.
- [22] O.P. Hauser, C. Hilbe, K. Chatterjee, and M.A. Nowak. Social dilemmas among unequals. *Nature*, 572:524–527, 2019.
- [23] J. Maynard Smith and G. R. Price. The logic of animal conflict. *Nature*, 246:15–18, 1973.
- [24] J. Maynard Smith. *Evolution and the Theory of Games*. Cambridge University Press, Cambridge, 1982.
- [25] J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge, UK, 1998.
- [26] M. A. Nowak. *Evolutionary dynamics*. Harvard University Press, Cambridge MA, 2006.
- [27] Tuomas W Sandholm and Robert H Crites. Multiagent reinforcement learning in the iterated prisoner’s dilemma. *BioScience*, 37:147–166, Jul 1996.
- [28] D. Fudenberg and D. Levine. *The Theory of Learning in Games*. MIT Press, Cambridge MA, 1998.
- [29] M. W. Macy and A. Flache. Learning dynamics in social dilemmas. *Proceedings of the National Academy of Sciences USA*, 99:7229–7236, 2002.
- [30] Marco Pangallo, James B.T. Sanders, Tobias Galla, and J. Doyne Farmer. Towards a taxonomy of learning dynamics in 2×2 games. *Games and Economic Behavior*, 132:1–21, 2022.

- [31] M. Broom, C. Cannings, and G.T. Vickers. Multi-player matrix games. *Bulletin of Mathematical Biology*, 59(5):931–952, 1997.
- [32] Maciej Bukowski and Jacek Miekisz. Evolutionary and asymptotic stability in symmetric multi-player games. *International Journal of Game Theory*, 33(1):41–54, Dec 2004.
- [33] P. D. Taylor and L. Jonker. Evolutionarily stable strategies and game dynamics. *Mathematical Biosciences*, 40:145–156, 1978.
- [34] C. Hauert, F. Michor, M. A. Nowak, and M. Doebeli. Synergy and discounting of cooperation in social dilemmas. *Journal of Theoretical Biology*, 239:195–202, 2006.
- [35] C. S. Gokhale and A. Traulsen. Evolutionary games in the multiverse. *Proceedings of the National Academy of Sciences USA*, 107:5500–5504, 2010.
- [36] Ross Cressman and Yi Tao. The replicator equation and other game dynamics. *Proceedings of the National Academy of Sciences USA*, 111:10810–10817, 2014.
- [37] J. Peña, L. Lehmann, and G. Nöldeke. Gains from switching and evolutionary stability in multi-player matrix games. *Journal of Theoretical Biology*, 346:23–33, 2014.
- [38] Manh Hong Duong and The Anh Han. On Equilibrium Properties of the Replicator–Mutator Equation in Deterministic and Random Games. *Dynamic Games and Applications*, 10(3):641–663, 2020.
- [39] Karl Tuyls, Julien Pérolat, Marc Lanctot, Georg Ostrovski, Rahul Savani, Joel Z. Leibo, Toby Ord, Thore Graepel, and Shane Legg. Symmetric Decomposition of Asymmetric Games. *Scientific Reports*, 8(1):1–20, 2018.
- [40] M. A. Nowak, A. Sasaki, C. Taylor, and D. Fudenberg. Emergence of cooperation and evolutionary stability in finite populations. *Nature*, 428:646–650, 2004.
- [41] A. Traulsen and C. Hauert. Stochastic evolutionary game dynamics. In Heinz Georg Schuster, editor, *Reviews of Nonlinear Dynamics and Complexity*, pages 25–61. Wiley-VCH, Weinheim, 2009.
- [42] T. Antal, M. A. Nowak, and A. Traulsen. Strategy abundance in 2×2 games for arbitrary mutation rates. *Journal of Theoretical Biology*, 257:340–344, 2009.
- [43] T. Antal, A. Traulsen, H. Ohtsuki, C. E. Tarnita, and M. A. Nowak. Mutation-selection equilibrium in games with multiple strategies. *Journal of Theoretical Biology*, 258:614–622, 2009.
- [44] T. Sekiguchi and H. Ohtsuki. Fixation probabilities of strategies for bimatrix games in finite populations. *Dynamic Games and Applications*, 7:93–111, 2017.

- [45] Bin Wu, Arne Traulsen, and Chaitanya S. Gokhale. Dynamic properties of evolutionary multi-player games in finite populations. *Games*, 4(2):182–199, 2013.
- [46] Karl Tuyls, Ann Nowe, Tom Lenaerts, and Bernard Manderick. An evolutionary game theoretic perspective on learning in multi-agent systems. In *Information, Interaction, and Agency*, pages 133–166. 2005.
- [47] Tobias Galla and J Doyne Farmer. Complex dynamics in learning complicated games. *Proceedings of the National Academy of Sciences USA*, 110(4):1232–1236, Jan 2013.
- [48] W. Barfuss, J. F. Donges, and J. Kurths. Deterministic limit of temporal difference reinforcement learning for stochastic games. *Physical Review E*, 99:043305, 2019.
- [49] Wolfram Barfuss, Jonathan F. Donges, Vítor V. Vasconcelos, Jürgen Kurths, and Simon A. Levin. Caring for the future can turn tragedy into comedy for long-term collective action under risk of collapse. *Proceedings of the National Academy of Sciences of the United States of America*, 117(23):12915–12922, 2020.
- [50] Marta C. Couto, Stefano Giaimo, and Christian Hilbe. Introspection dynamics: A simple model of counterfactual learning in asymmetric games. *New Journal of Physics*, 24(6):63010, 2022.