Introspection dynamics in asymmetric multiplayers games

Marta Couto[1*] and Saptarshi Pal[1*]

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

* co-first authors

Abstract

Abstract

Evolutionary game theory and models of learning provide a powerful framework to describe strategic decision-making in social interactions. In the simplest case, these models describe games among two identical players. However, many interactions in everyday life are more complex. They involve more than two players who may differ in their available actions and in their incentives to choose each action. Such interactions can be captured by asymmetric multiplayer games. Recently, introspection dynamics has been introduced to explore such asymmetric games. According to this dynamics, at each time step players compare their current strategy to an alternative strategy. If the alternative strategy results in a payoff advantage, it is more likely adopted. This model provides a simple way to compute the players' long-run probability of adopting each of their strategies. In this paper, we extend some of the previous results of introspection dynamics for 2-player asymmetric games to games with arbitrarily many players. First, we derive a formula that allows us to numerically compute the stationary distribution of introspection dynamics for any multiplayer asymmetric game. Second, we obtain explicit expressions of the stationary distribution for two special cases. These cases are additive games (where the payoff difference that a player gains by unilaterally switching to a different action is independent of the actions of their co-players), and symmetric multiplayer games with two strategies. To illustrate our results, we revisit several classical games such as the public goods game.

Keywords: multiplayer games, asymmetric games, non-linear interactions, introspection dynamics, strategy abundance, additive games.

Introduction

Social behavior has been studied extensively through pairwise interactions [1]. Despite their simplicity, they provide important insights, such as how populations can sustain cooperation [2–4]. Yet, many interesting collective behaviors occur when multiple individuals interact simultaneously [5–12]. Most of these situations cannot be captured by the sum of several pairwise interactions. Thus, to account for such non-linearities, one needs to consider multiplayer games [10]. A well-known effect that only emerges when more than two players are present is the "second-order free-riding problem" [13]. 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 [14]. 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 long as cooperation is incentivized by punishment (or rewards) of others, individuals can take advantage by defecting in this second-order dilemma. As such, this tempts everyone not to engage in costly punishment (or rewarding), and cooperation can break down. For this reason, peer-punishment and sanctioning institutions can be hard to sustain without additional mechanisms [15–19].

Another interesting effect that can be explored with multiplayer games is the scale or size of the interaction itself. In situations that require some sort of coordination and where expectations on others play an important role in one's decisions, a growing group size might hinder the optimal outcome [6]. Likewise, it has been shown that it is hard to cooperate in large groups [11, 20]. This is not a general effect, though. It can happen that adding more players actually promotes cooperation. Gokhale and Traulsen (2011) investigate the 2-player and 3-player cases of a task allocation problem [21]. Although the underlying rules defining the interactions are the same for the two cases, simply introducing a third player changes the overall game, resulting in a decrease in free-riding. Additionally, interaction group size can vary in a population of players. There, not only the average group size can have an important effect, but also the variance of the group size distribution [22, 23].

Complexity further increases when players differ significantly among themselves. This diversity can be captured by asymmetric games [24–32]. In symmetric games, all players are indistinguishable. Thus, to fully characterise the state of the game, we only require to know the number of players playing each strategy. Conversely, in asymmetric games, players can differ in their available actions and in their incentives to choose each action; players are of different types or roles. Therefore, they can have uneven effects on others' payoffs too. For example, in public goods games and collective-risk dilemmas, players can have different initial endowments (or wealth), productivitites, costs, risk perceptions, or risk exposures [32–37]. Hence, to fully describe the 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 evolutionary game theory (EGT) [1, 38–40] to learning models [41–44], game theoretic frameworks have been widely used to study strategic behavior. The concept of "evolutionary stable strategy" (ESS), originally proposed for parwise encounters [38], was extended for multiplayer games [5, 45, 46]. Also the well-known replicator equation [1, 47] easily comprises multiplayer games [22, 48–51]. More recently, the replicator-mutator equation was applied to study the dynamics of multiplayer games, too [52]. 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" [1, 53, 54]) or there is a single population of players where each can play the two types or roles ("role games") [1]. For role games, a strategy must include the action played in each of the two roles. The case of asymmetric games with more than two players is substantially less studied within deterministic EGT. Gokhale and Traulsen (2012) and Zhang et al. (2022) are two exceptions [53, 55]. Notably, although these works study multiplayer games, they consider, at most, two different types (drawn from two populations), which leaves out the exploration of full asymmetry.

Also stochastic evolutionary game dynamics [56, 57] provides several models for studying multiplayer and asymmetric games. Fixation probabilities [58] for asymmetric 2-player [59], asymmetric 3-player games [60], and symmetric multiplayer games [49, 61] were recently derived. Furthermore, average strategy abundances [62, 63] were obtained only for 2-player asymmetric games [29, 64] or multiplayer symmetric games [21, 65, 66]. For a review on evolutionary multiplayer games both in infinitely large populations as well as in finite populations, we refer to Gokhale and Traulsen (2014) [10].

Learning models (of strategic behavior) take a different approach from EGT [28, 41–44, 67–70]. There is no evolution of strategies in a population necessarily, but a process by which individuals learn strategies dynamically. Introspection dynamics has recently proven to be a useful learning model for tackling asymmetric games [32, 71–73]. This feature comes from the fact that players update their strategies by exploring their own set of strategies in the following simple way: each time, after a round of the game, a random 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, it is more likely adopted on the next round. We describe the model formally in the next section. While in Couto et al. (2022) [71] only 2-player games were considered, this framework is general enough to account for multiple players. Particularly, introspection dynamics allows a natural exploration of full asymmetry in many-player games compared to population models. For example, in imitation dynamics, one needs to specify who imitates who [34], as when players differ much, it might not make sense to assume they imitate any other player with the same likelihood. Introspection avoids this assumption because players' decisions only depend on their own payoffs.

Here, we extend previous results of pairwise games under introspection dynamics to multiplayer games. First, we derive a formula that allows us to numerically compute the stationary distribution of introspection dynamics for any multiplayer asymmetric game. Second, we obtain explicit expressions of the stationary distribution for two special cases. These cases are additive games (where the payoff difference that a player gains by unilaterally switching to a different action is independent of the actions of their co-players), and symmetric multiplayer games with two strategies. To illustrate our theoretical results, we analyse various multiplayer asymmetric social dilemmas, extending the framework in [48] to asymmetric games. Finally, we also study the asymmetric version of a public goods game with a rewarding stage [19].

Model of introspection dynamics in multiplayer games

We consider a normal form game with N players where $N \geq 2$. In the game, a player, say player i, can play actions from their action set, $\mathbf{A}_i := \{a_{i,1}, a_{i,2}, ..., a_{i,m_i}\}$. The action set of player i has m_i actions. In this model, players only use pure strategies. Therefore, there are only finitely many states of the game. More precisely, there are exactly $m_1 \times m_2 \times ... \times m_N$ states. We denote a state of the game by collecting the actions of all the players in the game in a vector, $\mathbf{a} := (a_1, a_2, ..., a_N)$ where $\mathbf{a} \in \mathbf{A} := \mathbf{A}_1 \times \mathbf{A}_2 \times ... \times \mathbf{A}_N$ and $a_i \in \mathbf{A}_i$. We also use the notation, $\mathbf{a} := (a_i, \mathbf{a}_{-i})$ to denote the state from the perspective of player i. In the state $(a_i, \mathbf{a}_{-i})$, player i plays the action $a_i \in \mathbf{A}_i$ and their co-players play the action $\mathbf{a}_{-i} \in \mathbf{A}_{-i}$ where $\mathbf{A}_{-i}$ is defined as $\mathbf{A}_{-i} := \prod_{j \neq i} \mathbf{A}_j$. In this paper, we use bold font letters to denote vectors and matrices. We use the corresponding normal font letters with subscripts to denote elements of the vectors (or matrices). 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 $\pi_i(a_i, \mathbf{a}_{-i})$.

Since players only use pure strategies in this model, we use the terms strategies and actions interchangeably throughout the whole paper. In this model, players update their strategies over time using the introspection dynamics [71]. At every time step, one randomly chosen player can update their strategy. The randomly chosen player, say i, currently playing action $a_{i,k}$, compares their current payoff to the payoff that they would obtain if they played a randomly selected action, $a_{i,l} \neq a_{i,k}$, from their action set $\mathbf{A}_i$. This comparison is done while assuming that the co-players do not change their respective actions. When the co-players of player i play $\mathbf{a}_{-i}$, player i changes from action $a_{i,k}$ to the new action $a_{i,l}$ with the probability,

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

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 actions. At $\beta = 0$, players update to a randomly chosen strategy with probablity 0.5. For $\beta > 0$, players update to the alternative strategy under consideration with probablity greater than 0.5 (or less than 0.5) if the switch gives them a non-zero increase (or decrease) in the payoffs.

Introspection dynamics can be studied by analyzing properties of the transition matrix, $\mathbf{T}$ of the resulting dynamical process. The transition matrix element $\mathrm{T}_{\mathbf{a},\mathbf{b}}$ 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. In order to formally define the transition matrix, we first need to introduce some notations and definitions. We start by defining the neighbourhood set of $\mathbf{a}$ as,

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

$$\mathrm{Neb}(\mathbf{a}) := \{\mathbf{b} \in \mathbf{A} \big| \quad \exists j : b_j \neq a_j \wedge \mathbf{b}_{-j} = \mathbf{a}_{-j}\} \tag{2}$$

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

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

$$a_{\mathrm{I}(\mathbf{a},\mathbf{b})} \neq b_{\mathrm{I}(\mathbf{a},\mathbf{b})} \tag{3}$$

In the previous example, the index of difference between the neighbouring states $(\mathrm{C}, \mathrm{C}, \mathrm{C})$ and $(\mathrm{C}, \mathrm{C}, \mathrm{D})$ is 3. Using the above definitions, one can formally define the transition matrix of introspection dynamics with:

$$\mathrm{T}_{\mathbf{a},\mathbf{b}} = \begin{cases} \frac{1}{N(m_j - 1)} \cdot p_{a_j \to b_j}(\mathbf{a}_{-j}) & \text{if } \mathbf{b} \in \mathrm{Neb}(\mathbf{a}) \quad \text{and,} \quad j = \mathrm{I}(\mathbf{a}, \mathbf{b}) \\ \\ 0 & \text{if } \mathbf{b} \notin \mathrm{Neb}(\mathbf{a}) \\ \\ 1 - \sum_{\mathbf{c} \neq \mathbf{b}} \mathrm{T}_{\mathbf{a},\mathbf{c}} & \text{if } \mathbf{a} = \mathbf{b} \end{cases} \tag{4}$$

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. We introduce a sufficient condition for the stationary distribution of $\mathbf{T}$ to be unique.

When the selection strength, β is finite, the transition matrix of introspection dynamics has a unique stationary distribution. 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) and there are only finitely many states of the game, every state is reachable in a finite number of steps from any 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 $\mathbf{u} := (\mathrm{u}_{\mathbf{a}})_{\mathbf{a}\in\mathbf{A}}$ which satisfies the conditions:

$$\mathbf{uT} = \mathbf{u} \tag{5}$$

$$\mathbf{u1} = 1 \tag{6}$$

where $\mathbf{1}$ is the column vector with size same as $\mathbf{u}$ and has all elements as 1. For all the analytical results in this paper, we consider β to be finite so that stationary distributions of the processes are unique.

The above equations only present an implicit representation of the stationary distribution $\mathbf{u}$. The stationary distribution can be explictly calculated by the following expression (which is derived using Eq. (5) and (6)) as:

$$\mathbf{u} = \mathbf{1}^{\intercal}(\mathbb{1} + \mathbf{U} - \mathbf{T})^{-1} \tag{7}$$

where $\mathbf{U}$ is a square matrix of the same size as $\mathbf{T}$ with all elements equal to 1 and $\mathbb{1}$ is the identity matrix. The matrix $\mathbb{1} + \mathbf{U} - \mathbf{T}$ is invertible when $\mathbf{T}$ is a primitive matrix [71]. Using Eq. (7) one can compute the unique stationary distribution of (a finite β-) introspection dynamics for any normal form game (with arbitrary number of asymmetric players).

The stationary distribution element $\mathrm{u}_{\mathbf{a}}$ 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. That is, the probability that player i plays action $a \in \mathbf{A}_i$ in the long run, $\xi_{i,a}$, can be computed as,

$$\xi_{i,a} := \sum_{\mathbf{q}\in\mathbf{A}_{-i}} \mathrm{u}_{(a,\mathbf{q})} \tag{8}$$

Additive games and their properties under introspection dynamics

In this section we discuss the stationary properties of introspection dynamics when players learn to play strategies in a special class of games: additive games [30, 51]. 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 $\mathbf{q} \in \mathbf{A}_{-i}$,

$$\pi_i(x, \mathbf{q}) - \pi_i(y, \mathbf{q}) =: f_i(x, y) \tag{9}$$

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* [51]. For games with this property, the stationary distribution of introspection dynamics takes a simple form,

Proposition 1. *When β is finite, the unique stationary distribution, $\mathbf{u} = (\mathrm{u}_{\mathbf{a}})_{\mathbf{a} \in \mathbf{A}}$, of introspection dynamics for the N-player additive game is given by:*

$$\mathrm{u}_{\mathbf{a}} = \prod_{j=1}^{N} \frac{1}{\sum\limits_{a' \in \mathbf{A}_j} e^{\beta f_j(a', a_j)}} \tag{10}$$

where, $f_j(a', a_j)$ is the co-player independent payoff difference given by Eq. (9).

For a proof of the Propositions and Corollaries, please see Appendix. Using the stationary distribution and Eq. (8), one can also exactly compute the cumulative probabilities with which players play their actions in the long run (i.e., the marginal distributions). In this regard, introspection learning in additive games is particularly interesting. The stationary distribution and the marginal distributions of introspection dynamics in additive games are related in a special way,

Proposition 2. *Let $\mathbf{u} = (\mathrm{u}_{\mathbf{a}})_{\mathbf{a} \in \mathbf{A}}$ be the unique stationary distribution of introspection dynamics with finite β for the N-player additive game. Then, $\mathrm{u}_{\mathbf{a}}$ is the product of the marginal probabilities that each player plays their respective actions in $\mathbf{a}$. That is,*

$$\mathrm{u}_{\mathbf{a}} = \prod_{j=1}^{N} \xi_{j,a_j} \tag{11}$$

For the N-player additive game, ξ_{j,a_j} is given by,

$$\xi_{j,a_j} = \frac{1}{\sum\limits_{a' \in \mathbf{A}_j} e^{\beta f_j(a', a_j)}} \tag{12}$$

where, $f_j(a', a_j)$ is the co-player independent payoff difference given by Eq. (9).

The above proposition states that for additive games, the stationary distribution of introspection dynamics can be factorized into its corresponding marginals. In the long run, the probability that players play the state $\mathbf{a} = (a_1, a_2, ..., a_N)$ is the product of the cumulative probabilities that player 1 plays a_1, player 2 plays a_2 and so on. This property of the additive game was already shown for the simple two-player, two-action donation game in Couto et al. [71]. Here we extend that result for any additive game with arbitrary number of players, each having an arbitrary number of strategies. In the next section we use the well-studied example of the linear public goods game (an additive game) to illustrate these results.

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 by c_i and b_i respectively. We define an indicator function $\alpha(.)$ to map the action of cooperation to 1 and the action of defection to 0. That is $\alpha(\mathrm{C}) = 1$ and $\alpha(\mathrm{D}) = 0$. The payoff of player i when the state of the game is $\mathbf{a}$ is given by:

$$\pi_i(\mathbf{a}) = \frac{1}{N} \sum_{j=1}^{N} \alpha(a_j) b_j - \alpha(a_i) c_i \tag{13}$$

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,

$$\pi_i(\mathrm{D}, \mathbf{q}) - \pi_i(\mathrm{C}, \mathbf{q}) = c_i - \frac{b_i}{N} =: f_i(\mathrm{D}, \mathrm{C}) \tag{14}$$

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 easily identifiable dominated strategies. Defection dominates cooperation when $c_i > b_i/N$ while cooperation dominates defection when $c_i < b_i/N$. Using Proposition 1, one can derive the closed form expression for the stationary distribution of a $N-$player linear public goods game with two strategies.

Proposition 3. *When β is finite, the unique stationary distribution of introspection dynamics for a $N-$player linear public goods game is given by:*

$$\mathrm{u}_{\mathbf{a}} = \prod_{j=1}^{N} \frac{1}{1 + e^{sign(a_j)\beta f_j(\mathrm{D},\mathrm{C})}} \tag{15}$$

where,

$$sign(a) = \begin{cases} \quad 1 \quad \textit{if} \quad a = \mathrm{C} \\ - \quad 1 \quad \textit{if} \quad a = \mathrm{D} \end{cases} \tag{16}$$

We use a simple example to illustrate 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 cumulative probability that player 1 (or 2) cooperates and player 3 defects are $\xi_{1,\mathrm{C}} = \xi_{2,\mathrm{C}} = 0.417$ and $\xi_{3,\mathrm{D}} = 0.697$ respectively. With the exact values, one can confirm the factorizing property of the stationary distribution for additive games in this example (i.e., Proposition 2). That is, $\mathrm{u}_{\mathrm{CCD}} = 0.121 = \xi_{1,\mathrm{C}} \cdot \xi_{2,\mathrm{C}} \cdot \xi_{3,\mathrm{D}}$.

We use the Eq. (15) to analyze the stationary behaviour of players after a long run of introspection in a linear public goods game (LPGG). First, we study the simplest case where all players are symmetric (the cost and benefit for all the 4 players are c and 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 cooperators). When the parameters of the game are such that defection dominates cooperation ($b = 2, c = 1$, Fig. 1a), the stationary distribution of the process at high β indicates that in the long-run, states with higher number of cooperators are less likely than states with lower number of cooperators. However, for intermediate and low β, stationary results are qualitatively different. 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. 1b). Since every possible state is equiprobable in the limit of $\beta \to 0$, the outcome with 2 cooperators is most likely only because there are more states with 2 cooperators than states with any other number of cooperators.

Naturally, β plays an important role in determining the overall cooperation in the long run. When β is low, average cooperation varies weakly with the strength of the dilemma, $b/N - c$ (Fig. 1c). Even when the temptation to defect is high ($b/N - c = -2$), players cooperate with a non-zero probability. Similarly, when cooperation is highly beneficial and strictly dominates defection ($b/N - c = 2$), players defect sometimes. At higher values of β, the stationary behaviour of players is more responsive to the payoffs and thus reflects an abrupt change near the parameters where the game transitions from defection-dominating to cooperation-dominating ($b/N - c = 0$).

To study what effects might appear due to asymmetry in the LPGG, we consider the game with 3 asymmetric players. All the players can 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. The terms δ_b and δ_c represents the strength of asymmetry between the three players (a higher absolute value of δ indicating a bigger asymmetry). When the players only differ in their cost of cooperation ($\delta_b = 0$ and $\delta_c = 0.5$, Fig 2a, left), their relative cooperation in the long run reflects their relative ability to cooperate. The player with the lowest cooperation cost (player 3), cooperates with the highest probability (and *vice-versa*, Fig 2a, right). Similarly, when players only differ in their ability to produce the public good ($\delta_b = 1$ and $\delta_c = 0$, Fig 2b left), their relative cooperation in the long run reflects the relative benefits they provide with their cooperation (Fig 2b, right). In this example, if we consider that the reference player provides a benefit of 2 units and has a cost of 1 unit (in which case, defection always dominates cooperation for them), defection dominates cooperation for player 1 if and only if $\delta_b < 1 + 3\delta_c$ and for player 3 only when $\delta_b > 3\delta_c - 1$. These regions in the $\delta_b - \delta_c$ parameter plane that correspond to defection dominating cooperation are circumscribed by white dashed lines in Fig. 2c. When players learn to play at high selection strength, β, their cooperation frequency in the long-run reflect the rational play (Fig. 2c). 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 is only assured if players are aligned in their asymmetries (i.e., either $\delta_b < 3\delta_c + 1$ or $\delta_b > 3\delta_c - 1$). Or, in other words, if the player that has low cost of cooperation also provides a high benefit upon contribution, then cooperation is high in the long-run.

Games with two actions and their properties under introspection dynamics

In the previous section we studied the properties of additive games under introspection dynamics. In this section, we study the stationary properties of games that are a) not necessarily additive and b) have only two actions for each player. First, we study the symmetric version of such a game. A N-player symmetric normal form game with two actions has the following properties:

1. All players have the same action set $\mathcal{A}$. That is, $\mathbf{A}_1 = \mathbf{A}_2 = ... = \mathbf{A}_N := \mathcal{A}$. We denote this set by, $\mathcal{A} := \{\mathrm{C}, \mathrm{D}\}$.

2. Players have the same payoff when they play against the same composition of co-players. That is, for any $i, j \in \{1, 2, ..., N\}$, $a \in \mathcal{A}$ and $\mathbf{b} \in \mathcal{A}^{N-1}$,

$$\pi_i(a, \mathbf{b}) = \pi_j(a, \mathbf{b}) \tag{17}$$

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 co-players playing C by

$\pi^{\mathrm{C}}(j)$ and $\pi^{\mathrm{D}}(j)$ respectively. We denote with $f(j)$ the payoff difference earned by switching from D to C when there are j co-players playing C,

$$f(j) := \pi^{\mathrm{D}}(j) - \pi^{\mathrm{C}}(j) \tag{18}$$

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

Proposition 4. *When β is finite, the unique stationary distribution of introspection dynamics for the $N-$player symmetric normal form game with two actions, $\mathcal{A} = \{\mathrm{C}, \mathrm{D}\}$, $(\mathrm{u}_{\mathbf{a}})_{\mathbf{a}\in\mathcal{A}^N}$, is given by:*

$$\mathrm{u}_{\mathbf{a}} = \frac{1}{\Gamma}\prod_{j=1}^{\mathcal{C}(\mathbf{a})} e^{-\beta f(j-1)} \tag{19}$$

where $f(j)$ is defined as in Eq. (18) and $\mathcal{C}(\mathbf{a})$ is the number of cooperators in state $\mathbf{a}$. The term Γ is the normalization factor given by:

$$\Gamma = \sum_{\mathbf{a}'\in\mathcal{A}^N}\prod_{j=1}^{\mathcal{C}(\mathbf{a}')} e^{-\beta f(j-1)} \tag{20}$$

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 4 can be simply reformulated by relabelling the states as follows,

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

$$\mathrm{u}_k = \frac{1}{\Gamma}\cdot\binom{N}{k}\cdot\prod_{j=1}^{k} e^{-\beta f(j-1)} \tag{21}$$

where, k represents the number of C players in the state and $f(j)$ is defined as in Eq. (18). The term Γ is the normalization factor, given by,

$$\Gamma = \sum_{k=0}^{N}\binom{N}{k}\cdot\prod_{j=1}^{k} e^{-\beta f(j-1)} \tag{22}$$

The above corollary follows directly from Proposition 4. 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 binomal coefficient $\binom{N}{k}$. In the next section, we use the example of a non-linear public goods game to illustrate these results.

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. [48]. In the original paper, the authors propose a normal form game with symmetric players. The game's properties are determined by a parameter w that determines the nature of the public good. The players have two actions: cooperation, C and defection, D. Here, we extend their framework to account for players with asymmetric payoffs. Before we explain the asymmetric setup, we describe the original model briefly. In the symmetric case, all N players have the same cost of cooperation, c and they all generate the same benefit b for the public good. Unlike the linear public goods game, contributions to the public good are scaled by a factor that is determined by w and the number of cooperators in the group. The payoff of a defector and a cooperator in a group with k cooperators and $N-k$ defectors is given by,

$$\pi_i^{\mathrm{D}}(k) = \frac{b}{N}(1 + w + w^2 + ... + w^{k-1}) \tag{23}$$

$$\pi_i^{\mathrm{C}}(k) = \pi_i^{\mathrm{D}}(k) - c \tag{24}$$

The parameter w represents the non-linearity of the public good. 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 contribution of every new cooperator goes down by a factor, w (compared to the last cooperator). The public goods is said to be discounting in this case. On the other hand when $w > 1$, every new contributor is more valuable than the previous one. The public good is said to be synergistic in this case. For the symmetric case, the relationship between the cost to benefit ratio, cN/b, and the discount/synergy factor, w, determines the type of social dilemma arising from the game. In principle, this framework can produce generalizations of the prisoner's dilemma (D dominating C), the snowdrift game (coexistence between C and D), the stag-hunt game (no dominance but existence of an internal unstable equilibrium) and the harmony game (C dominating D) with respect to its evolutionary trajectories under the replicator dynamics. For more details see Hauert et al. [48].

Now, we describe our extension of the original model to account for asymmetric players. Here, for player i, the cost of cooperation is c_i. The benefit that they can generate for the public good is b_i. 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 (just like the original model). However, now, since players are asymmetric it is not entirely clear in which order the contributions of cooperators should be discounted (or synergized). For example, consider that there are three cooperators in the group: player p, q and r. 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_p, b_q and b_r. In this model, we assume that all such permutations are equally likely, and therefore, the expected benefit provided by all three of them is given by $\bar{b}(1+w+w^2)$ where $\bar{b} = (b_p + b_q + b_r)/3$.

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

$$\pi_i(\mathrm{D}, \mathbf{a}_{-i}) = \begin{cases} \displaystyle\sum_{i=1}^{N} b_i \alpha(a_i) \cdot \frac{1}{N \cdot \mathcal{C}(\mathrm{D}, \mathbf{a}_{-i})} \cdot \Big(1 + w + w^2 + ...w^{\mathcal{C}(\mathrm{D}, \mathbf{a}_{-i})-1}\Big) & \text{if } \mathcal{C}(\mathrm{D}, \mathbf{a}_{-i}) \neq 0 \\ 0 & \text{if } \mathcal{C}(\mathrm{D}, \mathbf{a}_{-i}) = 0 \end{cases} \tag{25}$$

$$\pi_i(\mathrm{C}, \mathbf{a}_{-i}) = \sum_{i=1}^{N} b_i \alpha(a_i) \cdot \frac{1}{N \cdot \mathcal{C}(\mathrm{C}, \mathbf{a}_{-i})} \cdot \Big(1 + w + w^2 + ...w^{\mathcal{C}(\mathrm{C}, \mathbf{a}_{-i})-1}\Big) - c_i \tag{26}$$

where $\mathcal{C}(a, \mathbf{a}_{-i})$ counts the number of cooperators in state $(a, \mathbf{a}_{-i})$ and $\alpha(.)$ maps the actions C and D to 1 and 0 respectively. Note that the number of cooperators in the two states are related as: $\mathcal{C}(\mathrm{D}, \mathbf{a}_{-i}) = \mathcal{C}(\mathrm{C}, \mathbf{a}_{-i}) - 1$. We are interested in studying the long term stationary behaviour of players in this game when they learn through introspection. We first discuss results from the symmetric public goods game and then discuss results for the game with asymmetric players.

To compute the stationary distribution of introspection dynamics in this game, we use Eq. (21). In our symmetric example, we consider that every player in a $N-$player game can generate a benefit b of value 2. Before exploring the $c - w - N$ parameter space, we study four specific cases (with a 4 player game). In two of these cases, the public goods is discounted ($w = 0.5$, Fig. 3a left panels) and in two other cases, the public goods is synergistic ($w = 1.5$, Fig. 3a right panels). For each case, we consider two sub-cases: first, in which cost is high ($c = 1$, Fig. 3a top panels) and second, when cost is low ($c = 0.2$, Fig. 3a bottom panels). The four parameter combinations are chosen such that each of them corresponds to a unique social dilemma under the replicator dynamics. When selection strength is intermediate ($\beta = 5$), players sometimes play actions that are not optimal for the dilemma. For example, even when the parameters of the game make cooperation to be the dominated strategy ($w = 0.5, c = 1$), there is a single cooperator in the group in 20 % of the cases. When the parameters of the game reflect the stag-hunt dilemma ($c = 1, w = 1.5$), players are more likely to coordinate their actions in the long run. In the long-run, the probabilities that the whole group plays C or D is higher than the probabilities that there is a group with a mixture of C and D players. In contrast, when the parameters reflect the snowdrift game ($w = 0.5, c = 0.5$), we get the opposite effect. In the long run, mixed groups are more likely than homogeneous groups. Finally, when the parameters of the game make defection the dominated action

($w = 1.5, c = 0.2$), all players learn to cooperate 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 3b. First, let us consider the case when the group size is fixed at 4 players (the $c-w$ plane in Fig3b). In that case, if the cost of cooperation is restrictively high, the average cooperation rate is negligible and does not change with the change in the nature of the public good. In contrary, when the cost is not restrictively high, the discount/synergy parameter, w, determines the frequency with which players cooperate in the long run. A higher w for the public good would result in higher cooperation (and *vice-versa*). Next, we consider the case where the cost of cooperation is fixed (the $N - w$ plane in Fig. 3b). The cost is fixed to a value such that in a synergistic PG ($w > 1$), the cooperation frequency is almost 1 in the long run for any group size. 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. In discounted public goods, cooperation rates fall with increase in group sizes.

We also study introspection dynamics in this game with asymmetric players. We use the same setup that we used for studying the asymmetric linear public goods. The average frequency of cooperation per player is summarized in Supplementary Figures 1 and 2. In Supplementary Figure 1, we study two cases, first in which the public good is synergistic and players have a high average cost, and second in which public good is discounted and players have a lower average cost. In both of these cases, players cooperate highly when they simultaneously have low cost and high benefit. The only noticebale difference between the two cases is the minimum relationship between the asymmetries δ_b and δ_c that results in high cooperation for the player with low cost and high benefit. When we observe individual cooperation frequency versus the synergy/discount factor, w (Supplementary Figure 2), we find that when players are symmetric with respect to just benefits (or just costs), the one with the lowest cost (or highest benefit) cooperates with a high probability across all types of public goods, even for a high value of average cost.

Application: Introspection learning in a game with cooperation and rewards

In all the examples that we have studied so far, players can only choose between two actions (pure strategies). Introspection dynamics is particularly useful when players can use larger strategy sets. In this section, we study the stationary behaviour of players in a game where each player has 16 possible pure strategies. To this end, we adopt the multiplayer cooperation and rewarding game from Pal and Hilbe [19]. In this game, there are two stages: in stage 1, players decide whether or not they contribute to a linear public good and in stage 2, they decide whether or not they reward their peers. When a player contributes to the public good, they pay a cost c_i but generate a benefit worth $r_i c_i$ that is equally shared by everyone. When a player rewards a peer, they provide them a benefit of ρ while incurring the cost of rewarding, γ_i to self. In between the stages, players get full information about the contribution of their peers. In the rewarding stage, players have four possible strategies: they can either reward all the peers who contributed (social rewarding), reward all the peers who defected (antisocial rewarding), reward all peers irrespective of contribution (always rewarding) or reward none of the peers (never rewarding). Before stage 1 commences, player i knows with some probability, λ_i, the rewarding strategy of all their peers. In stage 1, players can have four possible strategies: they can either contribute or defect unconditionally or they can be conditional cooperators or conditional defectors. Conditional cooperators (or defectors) contribute (or do not contribute) when they have no information about their peers (which happens with probability $1 - \lambda_i$). When a conditional player, i, knows the rewarding strategy of all their peers (which happens with probability λ_i) and finds that there are n_{SR} social rewarders and n_{AR} antisocial rewarders among his peers, they cooperate if and only if the marginal gain from rewards for choosing cooperation over defection outweighs the effective cost of cooperation. That is,

$$\rho(n_{\mathrm{SR}} - n_{\mathrm{AR}}) \geq c_i \left(1 - \frac{r_i}{N}\right) \tag{27}$$

Combining the two stages, players can use one of 16 possible strategies (4 in stage 1 and 4 in stage 2). In the simple case where players are identical, one can characterize the Nash equilibria of the game and identify the conditions which allow an equilibrium where all players contribute in the first stage and reward peers in second stage [19]. In the symmetric case, full cooperation and rewarding is feasible in equilibrium when all players have sufficient information about each other and the reward benefit ρ is neither too high, nor too low. In this section, we study three simple cases of asymmetry between players to demonstrate how these asymmetric players may learn to play the game through introspection dynamics. The three specific examples that we show demonstrate that with introspection dynamics, asymmetric players can end up taking different roles in the long run to produce the public good. To this end, we consider a 3-player game in which player 1 and 2 are identical but player 3 is asymmetric to them in some aspect. We consider three cases. In each case the asymmetric player either has **a**) a higher cost of rewarding $\gamma_3 > \gamma_1$ or, **b**) low productivitiy $r_3 < r_1$ or, **c**) or, less information about peers $\lambda_3 < \lambda_1$ than their peers. We use Eq. (7) to exactly compute the expected abundances of the 16 strategies for each player.

In the case where player 3 is asymmetric with respect to their cost of rewarding, the long-run outcome of introspection reflects a division in labour between the players in producing the public good (Fig. 4a). The players, to whom rewarding is less costly (player 1 and player 2), reward cooperation with a higher probability than to whom rewarding is very costly (player 3). In return, player 3 learns to respond by contributing with more probability than their co-players. With these specific parameters, one player takes up the role of providing the highest per-capita contribution while the others compensate with costly rewarding. When the asymmetric player differs only in their productivity, a different effect may appear in the long run (Fig 4b). In this case, the less productive player free-rides on the cooperation of their higher productive peers, but eventually reward the cooperation of their peers nonetheless. The asymmetric player free-rides but does not second-order free ride. The probability with which the less productive player rewards others in the long run is slightly higher than the probability with which the contributing individuals reward each other. Finally, we consider the case where the asymmetric individual differs from others in terms of the information players have about others' rewarding strategy (Fig 4c). In this case, the asymmetric player knows others' strategy with a considerably less chance than their peers. In the long run, the asymmetric player cooperates less on average than their peers. This is because the asymmetric individual faces less instances where they can opportunistically cooperate with their co-players. However, both types of player reward cooperation almost equally and just enough to sustain cooperation.

Discussions and conclusion

We introduce introspection dynamics in N-player (a)symmetric games. In this learning model, at each time, one of the N players updates (or not) their strategy by comparing the payoffs of two strategies only: the one being currently played and a random prospective one. Clearly, this assumption implies a simple cognitive process. Players do not optimize over the entire set of strategies as, for example, in best-responde models [28, 74]. Furthermore, although conceptually similar, our model is also simpler than typical reinforcement learning models. For example, while we only have selection strength as a parameter (apart from payoffs), in Macy and Flache (2002) [43], there is a learning rate parameter (which could be comparable to our selection strength) but also an aspiration parameter which sets a payoff reference. In our model, the payoff reference is always the current one. All in all, whereas at each single time step, individuals are restricted to reason over two strategies only, as they iterate this step over time, they are able to fully explore the whole set of strategies, in a trial-and-error fashion.

Importantly, our model is also much simpler than the stochastic evolutionary game theory framework. While they both can involve solving the stationary distribution of a Markov process, they differ greatly in the state space size. Population models typically assume individuals play multiple games against (potentially all) other players in a population. As such, the state is defined by the number of players playing each strategy in the population(s). The number of states rapidly increases with the population size, the number of strategies, of players and of different types (in the case of asymmetric games). One can see how the mathematical analysis of multiplayer asymmetric games can become cumbersome. To deal with this issue, previous models frequently resorted to additional approximations, like low mutation rate [31, 75] and weak selection [76]. On the contrary, in introspection dynamics, the states of the Markov process correspond to the outcome of a single (focal) game: for a N-player game, where player i has m_i possible actions, there are $m_1 \times m_2 \times ... \times m_N$ states. This feature hugely reduces our state space size, which is key for obtaining exact results.

Here, we thus provide a general explicit formula (Eq. 7) that easily computes the stationary distribution of any multiplayer asymmetric game under introspection dynamics. Note that this formula is usefull for the exploration of many-strategy games in the full range of selection strenght. Additionally, we show that it is possible to obtain some analytical expressions for the long-run average strategy abundances. We start by analysing the set of additive games, for which the gain from switching between any two actions is constant, regardless of what co-players do. Due to this simple feature, additive games allow for the most general close-form expression for the stationary distribution (regarding the number of players, of strategies, and asymmetry of the game). Additionally, we find that for any additive game, the joint distribution of strategies factorizes over the marginal distribution of strategies. For more general games, we provide the stationary distribution formula for 2-strategy, symmetric games. Finally, we study several examples of social dilemmas. From those, we see that, despite the differences to other models pointed out above, we recover some previous qualitative results [48]. We also conlcude that players that have a lower cost or a higher benefit of cooperation learn to cooperate more frequently.

Introspection dynamics is a rather broad model. Here, we mainly focused on introducing a general framework. Still, we provide some examples to illustrate how it can be applied. Besides the generic public goods game, we study a 2-stage game, where players can choose among 16 strategies. There, individuals can reward their co-players condition on their previous cooperative (or not) behavior. Clearly, there are a number of ways in which our model can be further employed. For example, other researchers recently studied multiplayer games considering multiple games played concurrently [12], fluctuating environments [77], continuous strategies [78], or repeated interactions [11]. Also, a number of previous works considered complex population structures [65, 79–84]. As discussed above, introspection dynamics does not consider a population of players, making it simple to work with. However, it could be equally applicable to population models. In that case, players would obtain average payoffs either from well-mixed or network-bounded interactions, as usual, but update their strategies introspectively.

Acknowledgements

This work was supported by the European Research Council Starting Grant 850529 (E-DIRECT) and by the Max Planck Society. We would like to thank Christian Hilbe and the members of the Research Group Dynamics of Social Behavior for valuable feedback.

Appendix: Proofs

Proof. **Proof of Proposition** 1

Since β is finite,the stationary distribution $\mathbf{u} = (\mathrm{u}_{\mathbf{a}})_{\mathbf{a}\in\mathbf{A}}$ of the process is unique. The stationary distribution also satisfies the equalities in Eq. (5) and (6). Before continuing through the remainder of the proof, we introduce some short-cut notation that we will be using:

$$\mathrm{I}_{\mathbf{b}} := \mathrm{I}(\mathbf{b}, \mathbf{a}), \quad \textit{iff} \quad \mathbf{b} \in \mathrm{Neb}(\mathbf{a}) \tag{28}$$

$$\tau_{j,a_j} := \frac{1}{\sum\limits_{a'\in\mathbf{A}_j} e^{\beta f_j(a',a_j)}} \tag{29}$$

In order to show that the candidate stationary distribution, as proposed in Eq. (10) is the stationary distribution of the process, we need to show that the following are true:

$$\mathrm{T}_{\mathbf{a},\mathbf{a}}\mathrm{u}_{\mathbf{a}} + \sum_{\mathbf{b}\neq\mathbf{a}} \mathrm{T}_{\mathbf{b},\mathbf{a}}\mathrm{u}_{\mathbf{b}} = \mathrm{u}_{\mathbf{a}} \quad \forall \mathbf{a} \in \mathbf{A} \tag{30}$$

$$\sum_{\mathbf{a}\in\mathbf{A}} \mathrm{u}_{\mathbf{a}} = 1 \tag{31}$$

Using our short-cut notation τ and the expression for our candidate stationary distribution in Eq. (10), we can express the stationary distribution as:

$$\mathrm{u}_{\mathbf{a}} = \prod_{j=1}^{N} \tau_{j,a_j} \tag{32}$$

Using this expression, the left hand side of Eq. (30) can be simplified further with the steps:

$$\mathrm{T}_{\mathbf{a},\mathbf{a}}\mathrm{u}_{\mathbf{a}} + \sum_{\mathbf{b}\neq\mathbf{a}} \mathrm{T}_{\mathbf{b},\mathbf{a}}\mathrm{u}_{\mathbf{b}} \tag{33}$$

$$= \left(1 - \frac{1}{N} \sum_{\mathbf{b}\in\mathrm{Neb}(\mathbf{a})} \frac{1}{m_{\mathrm{I}_{\mathbf{b}}} - 1} \cdot p_{a_{\mathrm{I}_{\mathbf{b}}} \to b_{\mathrm{I}_{\mathbf{b}}}}\right) \mathrm{u}_{\mathbf{a}} + \frac{1}{N} \sum_{\mathbf{b}\in\mathrm{Neb}(\mathbf{a})} \frac{1}{m_{\mathrm{I}_{\mathbf{b}}} - 1} \cdot p_{b_{\mathrm{I}_{\mathbf{b}}} \to a_{\mathrm{I}_{\mathbf{b}}}} \cdot \mathrm{u}_{\mathbf{b}} \tag{34}$$

$$= \mathrm{u}_{\mathbf{a}} + \frac{1}{N} \sum_{\mathbf{b}\in\mathrm{Neb}(\mathbf{a})} \left(\prod_{k\neq I_{\mathbf{b}}} \tau_{k,a_k}\right) \left(p_{b_{\mathrm{I}_{\mathbf{b}}} \to a_{\mathrm{I}_{\mathbf{b}}}} \cdot \tau_{\mathrm{I}_{\mathbf{b}},a_{\mathrm{I}_{\mathbf{b}}}} - p_{a_{\mathrm{I}_{\mathbf{b}}} \to b_{\mathrm{I}_{\mathbf{b}}}} \cdot \tau_{\mathrm{I}_{\mathbf{b}},b_{\mathrm{I}_{\mathbf{b}}}}\right) \cdot \left(\frac{1}{m_{\mathrm{I}_{\mathbf{b}}} - 1}\right) \tag{35}$$

For an additive game, the expressions for $p_{b_{\mathrm{I}_\mathbf{b}} \to a_{\mathrm{I}_\mathbf{b}}}$ and $p_{a_{\mathrm{I}_\mathbf{b}} \to b_{\mathrm{I}_\mathbf{b}}}$ can be simply written as:

$$p_{b_{\mathrm{I}_\mathbf{b}} \to a_{\mathrm{I}_\mathbf{b}}} = \frac{1}{1 + e^{\beta f_{\mathrm{I}_\mathbf{b}}(b_{\mathrm{I}_\mathbf{b}}, a_{\mathrm{I}_\mathbf{b}})}} \tag{36}$$

$$p_{a_{\mathrm{I}_\mathbf{b}} \to b_{\mathrm{I}_\mathbf{b}}} = \frac{1}{1 + e^{\beta f_{\mathrm{I}_\mathbf{b}}(a_{\mathrm{I}_\mathbf{b}}, b_{\mathrm{I}_\mathbf{b}})}} \tag{37}$$

Using the above expressions and the expression for τ in Eq. (29), it can be shown that:

$$\left(p_{b_{\mathrm{I}_\mathbf{b}} \to a_{\mathrm{I}_\mathbf{b}}} \cdot \tau_{\mathrm{I}_\mathbf{b}, a_{\mathrm{I}_\mathbf{b}}} - p_{a_{\mathrm{I}_\mathbf{b}} \to b_{\mathrm{I}_\mathbf{b}}} \cdot \tau_{\mathrm{I}_\mathbf{b}, b_{\mathrm{I}_\mathbf{b}}} \right) = 0 \tag{38}$$

After plugging the equality in Eq. (38) into Eq. (35), we see that the left hand side of Eq. (30) simplifies to $\mathrm{u}_\mathbf{a}$. Now, to complete the proof we must check if Eq. (31) holds for our candidate distribution. Summing up the elements of the stationary distribution $\mathrm{u}_\mathbf{a}$ for all states $\mathbf{a} \in \mathbf{A}$:

$$\sum_{\mathbf{a} \in \mathbf{A}} \mathrm{u}_\mathbf{a} = \sum_{\mathbf{a} \in \mathbf{A}} \prod_{k=1}^{N} \tau_{k, a_k} = \sum_{\mathbf{a} \in \mathbf{A}} \frac{\prod_{k=1}^{N} e^{\beta \pi_k(a_k, \mathbf{q}_{-k})}}{\prod_{k=1}^{N} \sum_{a' \in \mathbf{A}_k} e^{\beta \pi_k(a', \mathbf{q}_{-k})}} \tag{39}$$

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

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

$$= \left(\prod_{k=1}^{N} \left(e^{\beta \pi_k(a_{k,1}, \mathbf{q}_{-k})} + ... + e^{\beta \pi_k(a_{k,m_k}, \mathbf{q}_{-k})} \right) \right)^{-1} \cdot \left(\sum_{\mathbf{a} \in \mathbf{A}} \prod_{k=1}^{N} e^{\beta \pi_k(a_k, \mathbf{q}_{-k})} \right) \tag{41}$$

$$\tag{42}$$

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

$$\sum_{\mathbf{a}\in\mathbf{A}} \mathrm{u}_{\mathbf{a}} = \left(\prod_{k=1}^{N}\left(e^{\beta\pi_k(a_{k,1},\mathbf{q}_{-k})} + ... + e^{\beta\pi_k(a_{k,m_k},\mathbf{q}_{-k})}\right)\right)^{-1} \cdot \left(\sum_{\mathbf{a}\in\mathbf{A}}\prod_{k=1}^{N} e^{\beta\pi_k(a_k,\mathbf{q}_{-k})}\right) \tag{43}$$

$$= \left(\sum_{\mathbf{a}\in\mathbf{A}}\prod_{k=1}^{N} e^{\beta\pi_k(a_k,\mathbf{q}_{-k})}\right)^{-1}\left(\sum_{\mathbf{a}\in\mathbf{A}}\prod_{k=1}^{N} e^{\beta\pi_k(a_k,\mathbf{q}_{-k})}\right) = 1 \tag{44}$$

The step from Eq. (43) to Eq. (44) involves multiplying out all the sums of exponents (where each term in the sum of exponents corresponds to payoff that player k receives by playing their actions against co-player composition, $\mathbf{q}_{-k}$). 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.

□

Proof. **Proof of Proposition** 2

Just like the previous proof, $\mathbf{p}_{-1}, \mathbf{p}_{-2}, ..., \mathbf{p}_{-N}$ are any arbitrary tuples from $\mathbf{A}_{-1}, \mathbf{A}_{-2}, ..., \mathbf{A}_{-N}$ 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 $\mathbf{u} = (\mathrm{u}_{\mathbf{a}})_{\mathbf{a}\in\mathbf{A}}$ 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. (10). We use this expression to calculate the marginal distribution of actions played at a particular state $\mathbf{a}$, $(\xi_{j,a_j})_{j\in\{1,2,...,N\}}$.

$$\xi_{j,a_j} = \sum_{\mathbf{q}\in\mathbf{A}_{-j}} \mathrm{u}_{(a_j,\mathbf{q})} \tag{45}$$

$$= \sum_{\mathbf{q}\in\mathbf{A}_{-j}}\left(\sum_{a'\in\mathbf{A}_j} e^{\beta f_j(a',a_j)}\right)^{-1}\prod_{k\neq j}\left(\sum_{a'\in\mathbf{A}_k} e^{\beta f_k(a',q_k)}\right)^{-1} \tag{46}$$

$$= \left(\prod_{k=1}^{N}\sum_{a'\in\mathbf{A}_k} e^{\beta\pi_k(a',\mathbf{p}_{-k})}\right)^{-1} \cdot e^{\beta\pi_j(a_j,\mathbf{p}_{-j})} \cdot \left(\sum_{\mathbf{q}\in\mathbf{A}_{-j}}\prod_{k\neq j} e^{\beta\pi_k(q_k,l_{-k})}\right) \tag{47}$$

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

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

The interchange of the sum and the product between the expressions in Eq. (48) and Eq. (49) 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 q_k) against co-players that play $\mathbf{p}_{-k}$. This is the similar to the step between Eq. (43) and Eq. (44) in the proof of Proposition 2.

$$= \left(\sum_{a' \in \mathbf{A}_j} e^{\beta (\pi_j(a', \mathbf{p}_{-j}) - \pi_j(a_j, \mathbf{p}_{-j}))} \right)^{-1} \tag{50}$$

$$= \sum_{a' \in \mathbf{A}_j} e^{\beta f_j(a', a_j)} \tag{51}$$

Using the expression in Eq. (51), we can confirm that for additive games, the product of the marginals is the stationary distribution,

$$\prod_{j=1}^{N} \xi_{j, a_j} = \mathrm{u}_{\mathbf{a}} \tag{52}$$

□

Proof. **Proof of Proposition** 3

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 1. Here, we provide an independent proof. The idea behind this proof is identical to the proof of Proposition 1.

Again, 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:

$$\bar{a}_j := \{\mathrm{D}, \mathrm{C}\} \setminus \{a_j\} \tag{53}$$

$$p_j := \frac{1}{1 + e^{\beta f_j(\mathrm{D}, \mathrm{C})}} \tag{54}$$

In addition we introduce an indicator function $\alpha(.)$ which maps the action C to 1 and the action D to 0. That is $\alpha(\mathrm{C}) := 1$ and $\alpha(\mathrm{D}) := 0$. Using these notations and Eq. (1) and (14) and utilizing our shortcut notation from above, we can write the probability that a player j updates to a_j from $\bar{a}_j$ while their co-players play $\mathbf{a}_{-j}$ as:

$$p_{\bar{a}_j} \rightarrow {}_{a_j}(\mathbf{a}_{-j}) = p_j sign(a_j) + \alpha(\bar{a}_j) \tag{55}$$

The candidate stationary distribution $\mathbf{u}$ given in Eq. (15) can be written down using our short-cut notation as:

$$\mathrm{u}_{\mathbf{a}} = \prod_{k=1}^{N} p_k sign(a_k) + \alpha(\bar{a}_k) \tag{56}$$

This stationary distribution must satisfy the following properties, which are also given in Eq (5) and (6):

$$\mathrm{u}_{\mathbf{a}} = \mathrm{T}_{\mathbf{a},\mathbf{a}}\mathrm{u}_{\mathbf{a}} + \sum_{\mathbf{b}\neq\mathbf{a}} \mathrm{T}_{\mathbf{b},\mathbf{a}}\mathrm{u}_{\mathbf{b}} \tag{57}$$

$$\sum_{\mathbf{a}\in\mathbf{A}} \mathrm{u}_{\mathbf{a}} = 1 \tag{58}$$

Where, the terms in the right hand side of Eq. (57) can be simplified using Eq. (1) and (4) as follows:

$$\mathrm{T}_{\mathbf{a},\mathbf{a}} = 1 - \sum_{k=1}^{N} \mathrm{T}_{(a_k,\mathbf{a}_{-k}),(\bar{a}_k,\mathbf{a}_{-k})} = 1 - \frac{1}{N}\sum_{k=1}^{N} p_k sign(\bar{a}_k) + \alpha(a_k) \tag{59}$$

and additionally, using Eq. (56) the second term can be simplified too:

$$\sum_{\mathbf{b}\neq\mathbf{a}} \mathrm{T}_{\mathbf{b},\mathbf{a}}\mathrm{u}_{\mathbf{b}} = \sum_{k=1}^{N} \mathrm{T}_{(\bar{a}_k,\mathbf{a}_{-k}),(a_k,\mathbf{a}_{-k})}\mathrm{u}_{(\bar{a}_k,\mathbf{a}_{-k})} \tag{60}$$

$$= \frac{1}{N}\sum_{k=1}^{N} \left(p_k sign(a_k) + \alpha(\bar{a}_k)\right) \mathrm{u}_{(\bar{a}_k,\mathbf{a}_{-k})} \tag{61}$$

$$= \frac{\mathrm{u}_{\mathbf{a}}}{N}\sum_{k=1}^{N} p_k sign(\bar{a}_k) + \alpha(a_k) \tag{62}$$

Now, using Eq. (59) and (62) one can show that the right hand side of Eq. (57) is the element of the stationary distribution, corresponding to the state $\mathbf{a}$, u_a. Now, to complete the proof, we must show that Eq. (58) 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:

$$\sum_{\mathbf{a}\in\mathbf{A}} \mathrm{u}_{\mathbf{a}} = \sum_{\mathbf{a}\in\mathbf{A}} \prod_{k=1}^{N} p_k sign(a_k) + \alpha(\bar{a}_k) \tag{63}$$

$$= \sum_{\mathbf{a}\in\mathbf{A}_{-N}} (1 - p_N) \prod_{k=1}^{N-1} p_k sign(a_k) + \alpha(\bar{a}_k) + p_N \prod_{k=1}^{N-1} p_k sign(a_k) + \alpha(\bar{a}_k) \tag{64}$$

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

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

□

Proof. **Proof of Proposition** 4

By construction, the candidate stationary distribution given by Eq. (19) and Eq. (20) is a probability distribution since it satisfies the condition in Eq. (6) and for any state $\mathbf{a}$, $\mathrm{u}_{\mathbf{a}}$ is between 0 and 1. Again, since β is finite the process will have a unique stationary distribution. Again, to show that the candidate stationary distribution is the unique stationary distribution, we need to check if Eq. (5) holds. That is, the condition in Eq. (57) must hold for all states $\mathbf{a}$. We re-introduce some notations that we will use in this proof:

$$\bar{a}^j := \{\mathrm{D}, \mathrm{C}\} \setminus \{a_j\} \tag{66}$$

$$\alpha(a) := \begin{cases} 1 & \text{if} \quad a = \mathrm{C} \\ 0 & \text{if} \quad a = \mathrm{D} \end{cases} \tag{67}$$

$$\mathcal{C}(\mathbf{a}) = \sum_{j=1}^{N} \alpha(a_j) \tag{68}$$

For this process, since there are only two actions, the first term in the right hand side of Eq. (57) can be simplified as:

$$\mathrm{u}_{\mathbf{a}}\mathrm{T}_{\mathbf{a},\mathbf{a}} = \mathrm{u}_{\mathbf{a}} - \mathrm{u}_{\mathbf{a}} \sum_{k=1}^{N} \mathrm{T}_{(a_k,\mathbf{a}_{-k}),(\bar{a}_k,\mathbf{a}_{-k})} \tag{69}$$

$$= \mathrm{u}_{\mathbf{a}} - \frac{\mathrm{u}_{\mathbf{a}}}{N} \sum_{k=1}^{N} \frac{1}{1+e^{sign(\bar{a}_k)\beta f(N_k)}} \tag{70}$$

Where, the function $sign(.)$ is defined as in Eq. (16) 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,

$$N_k := \sum_{j \neq k} \alpha(a_j) \tag{71}$$

The second term in the right hand side of Eq. (57) can be simplified as,

$$\sum_{\mathbf{b}\neq\mathbf{a}} \mathrm{T}_{\mathbf{b},\mathbf{a}}\mathrm{u}_{\mathbf{b}} = \sum_{k=1}^{N} \mathrm{T}_{(\bar{a}_k,\mathbf{a}_{-k}),(a_k,\mathbf{a}_{-k})}\mathrm{u}_{(\bar{a}_k,\mathbf{a}_{-k})} \tag{72}$$

$$= \frac{1}{N\Gamma} \sum_{k=1}^{N} \mathrm{T}_{(\bar{a}_k,\mathbf{a}_{-k}),(a_k,\mathbf{a}_{-k})} \prod_{j=1}^{\mathcal{C}((\bar{a}_k,\mathbf{a}_{-k}))} e^{-\beta f(j-1)} \tag{73}$$

$$= \frac{1}{N\Gamma} \sum_{k=1}^{N} \mathrm{T}_{(\bar{a}_k,\mathbf{a}_{-k}),(a_k,\mathbf{a}_{-k})} \left(\prod_{j=1}^{N_k} e^{-\beta f(j-1)} \right) \cdot e^{-\beta\alpha(\bar{a}_k) f(-\alpha(a_k)+N_k})\tag{74}$$

From Eq. (73) to Eq. (74), 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 $(\bar{a}_k, \mathbf{a}_{-k})$ of $\mathbf{a}$. For simplicity, we represent $\mathrm{T}_{(\bar{a}_k,\mathbf{a}_{-k}),(a_k,\mathbf{a}_{-k})}$ with just $\mathbf{T}$ in the next steps. We continue the simplification of Eq. (74) in the next steps by introducing terms that cancel each other.

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

The newly introduced term in Eq. (75) can be taken inside the product. Note that this term is 1 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 $\mathrm{u}_{\mathbf{a}}$. That is,

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

$$= \frac{1}{N}\sum_{k=1}^{N}\mathbf{T}\cdot\left(\frac{1}{\Gamma}\prod_{j=1}^{\mathcal{C}(\mathbf{a})} e^{-\beta f(j-1)}\right)\cdot\frac{e^{-\beta\alpha(\bar{a}_k)f(-\alpha(a_k)+N_k)}}{e^{-\beta\alpha(a_k)f(-\alpha(\bar{a}_k)+N_k)}} \tag{77}$$

$$= \frac{1}{N}\sum_{k=1}^{N}\mathrm{T}_{(\bar{a}_k,\mathbf{a}_{-k}),(a_k,\mathbf{a}_{-k})}\cdot\mathrm{u}_{\mathbf{a}}\cdot\frac{e^{-\beta\alpha(\bar{a}_k)f(-\alpha(a_k)+N_k)}}{e^{-\beta\alpha(a_k)f(-\alpha(\bar{a}_k)+N_k)}} \tag{78}$$

The fraction inside the sum in Eq. (78) can be simplified using the $sign(.)$ function (in 16) leading to further simplification of Eq. (78):

$$\sum_{\mathbf{b}\neq\mathbf{a}} \mathrm{T}_{\mathbf{b},\mathbf{a}}\mathrm{u}_{\mathbf{b}} = \frac{1}{N}\sum_{k=1}^{N}\mathrm{T}_{(\bar{a}_k,\mathbf{a}_{-k}),(a_k,\mathbf{a}_{-k})}\cdot\mathrm{u}_{\mathbf{a}}\cdot e^{sign(a_k)\beta f(N_k)} \tag{79}$$

In Eq. (79) we can replace the element of the transition matrix $\mathrm{T}_{(\bar{a}_k,\mathbf{a}_{-k}),(a_k,\mathbf{a}_{-k})}$ with,

$$\mathrm{T}_{(\bar{a}_k,\mathbf{a}_{-k}),(a_k,\mathbf{a}_{-k})} = \frac{1}{1+e^{sign(a_k)\beta f(N_k)}} \tag{80}$$

Using the expression for the transition matrix element from Eq. (80) into Eq. (79) and by using Eq. (70), we can simplify further:

$$\sum_{\mathbf{b}\neq\mathbf{a}} \mathrm{T}_{\mathbf{b},\mathbf{a}}\mathrm{u}_{\mathbf{b}} = \frac{\mathrm{u}_{\mathbf{a}}}{N}\sum_{k=1}^{N}\frac{1}{1+e^{sign(a_k)\beta f(N_k)}}\cdot e^{sign(a_k)\beta f(N_k)} \tag{81}$$

$$= \frac{\mathrm{u}_{\mathbf{a}}}{N}\sum_{k=1}^{N}\frac{1}{1+e^{sign(\bar{a}_k)\beta f(N_k)}} \tag{82}$$

$$= \mathrm{u}_{\mathbf{a}} - \mathrm{u}_{\mathbf{a}}\mathrm{T}_{\mathbf{a},\mathbf{a}} \tag{83}$$

The final step in the previous simplification shows that Eq. (57) holds for any $\mathbf{a}\in\{\mathrm{C},\mathrm{D}\}^N$. Therefore, the candidate distribution we propose in Eq. (19) is the unique stationary distribution of the symmetric N-player game with two strategies.

□

Proof. **Proof of Corollary** 1

To show this result we count how many states are identical to a state $\mathbf{a} \in \{\mathrm{C}, \mathrm{D}\}^N$ 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. 19. 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., $\mathrm{u}_{\mathbf{a}'} = \mathrm{u}_{\mathbf{a}}$).

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, u_k, is,

$$\mathrm{u}_k = \sum_{\mathcal{C}(\mathbf{a})=k} \mathrm{u}_{\mathbf{a}} = \frac{1}{\Gamma} \binom{N}{k} \prod_{j=1}^{k} e^{-\beta f(j-1)} \tag{84}$$

Where the normalization factor Γ can also be simplified as:

$$\Gamma = \sum_{k=0}^{N} \binom{N}{k} \prod_{j=1}^{k} e^{-\beta f(j-1)} \tag{85}$$

□

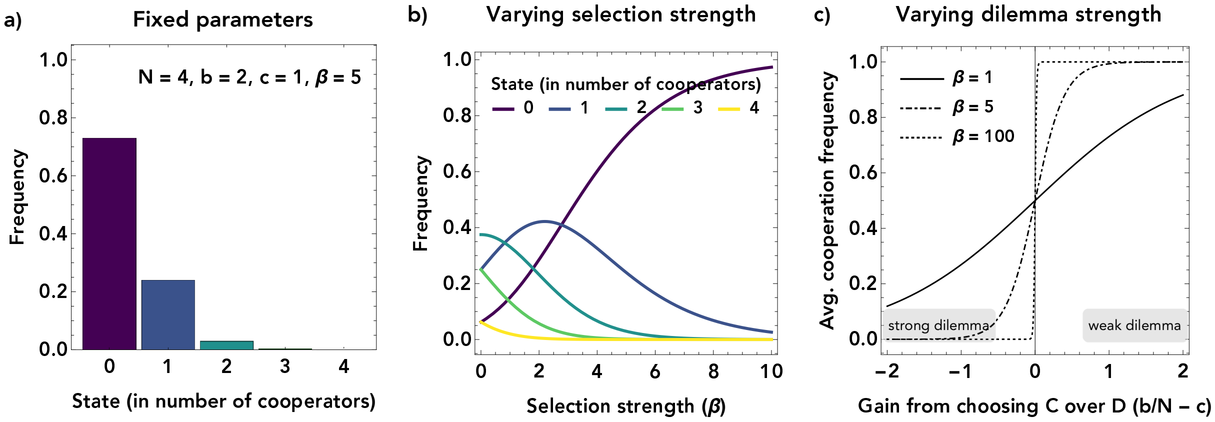

Figure 1: Introspection dynamics in a symmetric linear public goods game. The stationary distribution of introspection dynamics for a linear public goods game with all identical players is shown in this figure. For all the panels in this figure, we use the following parameters: $N = 4$ (group size), $b = 2$ (benefit provided to the public good upon cooperation), $c = 1$ (cost of cooperation) a) Here we show the frequency of each state in the stationary distribution of introspection dynamics. As players are identical, each state can be defined by the number of cooperators. We use a selection strength of $\beta = 5$ for the introspection dynamics. For this strength of selection, states with more cooperators are less likely than states with less cooperators in the stationary distribution b) Here we show the frequency of each state for varying selection strength, β. We use the same color code as the previous panel. Comparing neutrality ($\beta = 0$) with low to intermediate β values, we see that selection favors states other than 0 cooperators. Indeed, up to $\beta \approx 3$, state 0 is not the most frequent state in the long run. c) Average cooperation frequency for varying dilemma strength depends on the selection strength, β, and the parameters of the linear public goods game: b, c, N. We use the marginal gain of choosing cooperation over defection, $b/N - c$, as a measure of the dilemma strength. When this quantity is negative and low, we say that the dilemma is strong. In this case, chosing cooperation is strictly disadvantageous. When this quantity is positive and high, we say that the dilemma is weak. In this case, cooperation dominates defection. Typically, a linear public goods dilemma is defined to have a negative marginal gain. Here, we show the dilemma strength varying from -2 to 2. The results are shown for different values of selection strength, $\beta = 1, 5$ and 100. For high β, stationary distribution of introspection dynamics reflects the rational play. In the long-run player play the Nash equilibrium. When marginal gain is negative, defection is played with almost certainty (and *vice-versa*). For low β, however, we see that some cooperation is possible even when the dilemma is strong.

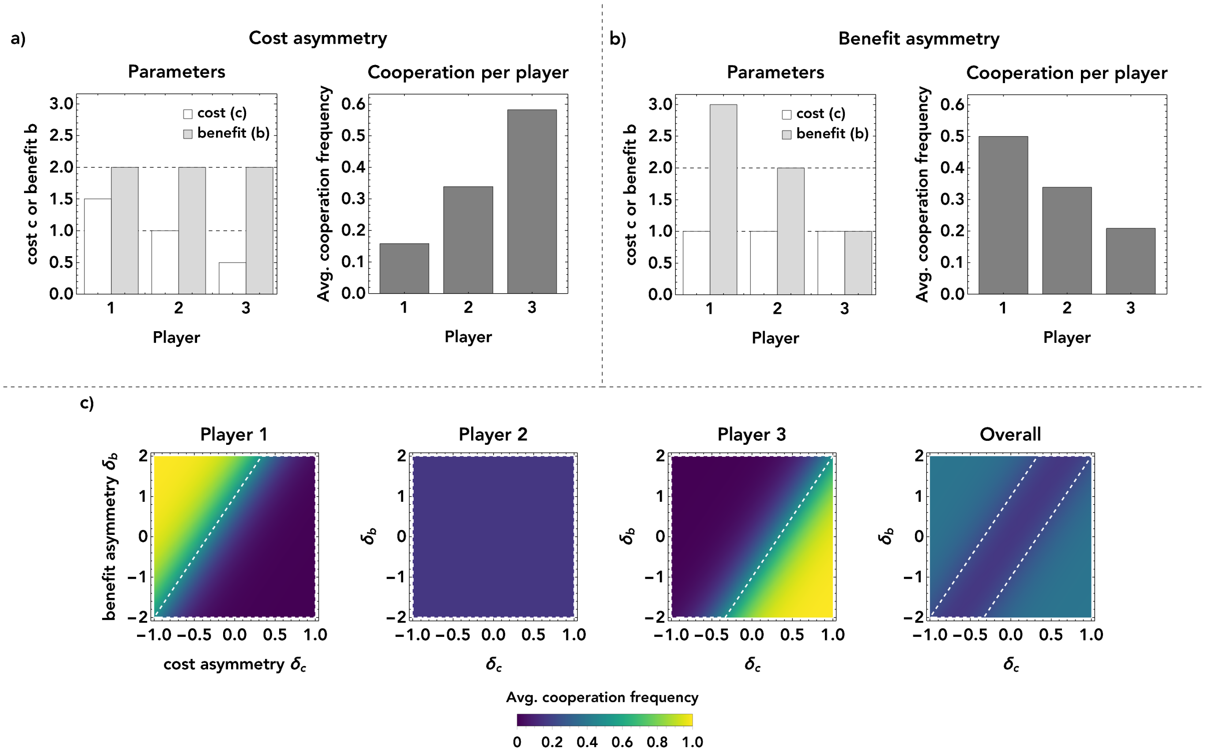

Figure 2: Introspection dynamics in an asymmetric linear public goods game. Three asymmetric players learn to play the linear public goods game through introspection dynamics. The cooperation probabilities in the stationary distribution are displayed here. For each of the upper panels (a and b), we show the cost of cooperation and the benefit provided upon cooperation for the players on the left and the average cooperation frequency in the long-run on the right. In this example, the cost of cooperation for players 1, 2 and 3 are $1 + \delta_c$, 1 and $1 - \delta_c$ respectively. The benefits that player 1, 2 and 3 provide upon cooperation are $2 + \delta_b$, 2 and $2 - \delta_b$ respectively. The cost and benefit for the reference player (player 2) are shown with black dashed lines in the left panels of a and b. In panel a, player 1 and 3 differ by 0.5 units in their cost of cooperation from the reference player. All the players provide the same benefit when they contribute (i.e., $\delta_b = 0$). Conversely, in panel b, benefit by player 1 and 3 differ from the reference player by 1 unit. The cost of cooperation is the same for all players ($c = 1$). In c), we vary the asymmetry strengths between the players, δ_c and δ_b, simultaneously and show both average individual cooperation frequency and the overall average cooperation frequency in the long-run. The reference player's cost and benefit are again 1 and 2 units respectively. The area within the white dashed lines represents the parameter values for which the marginal gain of choosing cooperation over defection is negative, for each single player and, in the right-most panel, for all players simultaneously. In this example, with this specific asymmetrical players, cooperation is only feasible in the long run if the asymmetries of players are aligned. That is, overall cooperation is high only when the individual with a low cost of cooperation has a high benefit value. For panels a and b we use a selection strength of $\beta = 2$ while for panel c, we use a selection strength of $\beta = 5$.

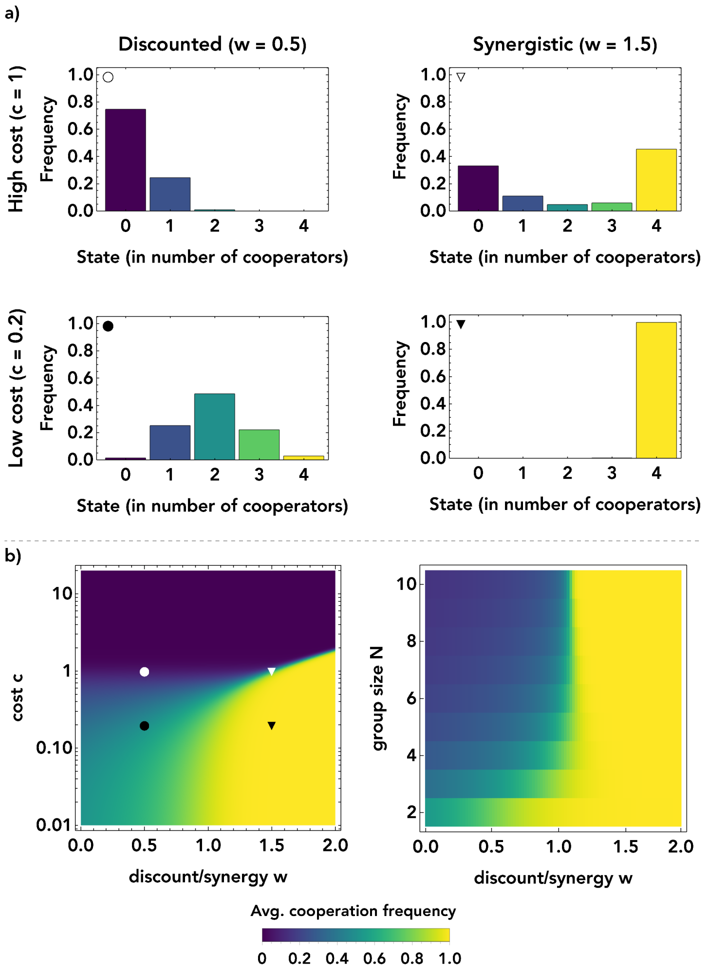

Figure 3: Introspection dynamics in a symmetric general public goods game. We study introspection dynamics in the general public goods game with 4 symmetric players, each having two possible actions - cooperation and defection. For a detailed description of the game, please see the main text. a) The frequency of each state in the stationary distribution of introspection dynamics in four types of multiplayer social dilemmas display qualitatively different results. The upper panels refer to a high cost of cooperation ($c = 1$) while the bottom panels to a low cost of cooperation ($c = 0.2$); left panels refer to a discounted public good ($w = 0.5$), and the right panels refer to a synergistic public good ($w = 1.5$). Each case is tagged with a symbol that places the particular case in the contour plot in panel b, b) Average cooperation frequency for varying discount/synergy factor, w, and varying cost of cooperation, c. Cooperation is feasible when costs are not restrictively high and the public good is not too discounted. c) Here we show the average cooperation frequency for varying discount/synergy factor, w, and group size N. For this plot, the cost of cooperation for each player is $c = 0.4$. The feasibility of cooperation drops with larger group sizes when the public good is discounted. For all panels, the benefit of cooperation generated by each player is worth 2 units. The selection strength of introspection dynamics is $\beta = 5$ for all panels.

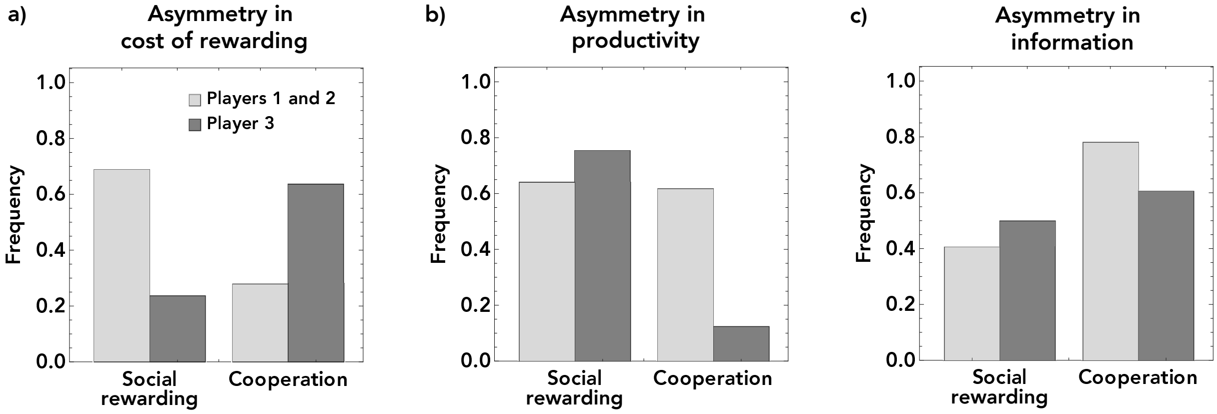

Figure 4: Intropection dynamics in the linear public goods game with peer rewarding. Here we study a game with three asymmetric players, each having 16 possible strategies. Players cooperate in a linear public goods and then reward each other in the next stage after everyone's contribution is revealed. In the first stage, players can condition their cooperation on the information they have about their co-players' rewarding strategies. For a full description of the model, please see the section on rewarding. In this example, players 1 and 2 are identical in all aspects while player 3 is asymmetric to them in only a single aspect. In here, we use Eq. (7) to plot the exact probability with which the asymmetric players cooperate and reward cooperation after a long run of introspection. We consider three types of asymmetry for player 3. a) First, we consider the case where player 3 has a high cost of rewarding compared to player 1 and 2, $0.7 = \gamma_3 > \gamma_1 = 0.1$. b) Then, we consider the case where player 3 is less productive than their co-players, $1.2 = r_3 < r_1 = 2$. c) Finally, we consider the case where player 3 has less information about co-players' rewarding strategies than their peers, that is, $0.1 = \lambda_3 < \lambda_1 = 0.9$. For all plots, we consider a high value for the selection strenght, $\beta = 10$. Unless otherwise mentioned, the following parameters are maintained for all panels: $c_i = 1$ (individual cost of cooperation), $r_i = 2$ (individual productivity), $\gamma_i = 0.1$ (individual cost of rewarding), $\lambda_i = 0.9$ (individual information about co-players' strategies). In panels a and b, the reward value ρ is 0.3 while for the last panel, c, the reward value $\rho = 1$.

References

[1] J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge, UK, 1998.

[2] R. Axelrod. *The evolution of cooperation*. Basic Books, New York, NY, 1984.

[3] Martin A Nowak. Five rules for the evolution of cooperation. *science*, 314(5805):1560–1563, 2006.

[4] M. A. Nowak and R. Highfield. *SuperCooperators: Altruism, Evolution, and Why We Need Each Other to Succeed*. Free Press, 2011.

[5] Günther Palm. Evolutionary stable strategies and game dynamics for n-person games. *Journal of Mathematical Biology*, 19(3):329–334, Jul 1984.

[6] B. Skyrms. *The Stag-Hunt Game and the Evolution of Social Structure*. Cambridge University Press, Cambridge, 2003.

[7] 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.

[8] 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.

[9] 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.

[10] C. S. Gokhale and A. Traulsen. Evolutionary multiplayer games. *Dynamic Games and Applications*, 4:468–488, 2014.

[11] 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.

[12] V. R. Venkateswaran and C. S. Gokhale. Evolutionary dynamics of complex multiple games. *Proceedings of the Royal Society B*, 286:20190900, 2019.

[13] J. H. Fowler. Altruistic punishment and the origin of cooperation. *Proceedings of the National Academy of Sciences USA*, 102(19):7047–7049, 2005.

[14] J. Henrich, R. McElreath, A. Barr, J. Ensminger, C. Barrett, A. Bolyanatz, J. C. Cardenas, M. Gurven, E. Gwako, N. Henrich, C. Lesorogol, F. Marlowe, D. Tracer, and J. Ziker. Costly punishment across human societies. *Science*, 312:1767–1770, 2006.

[15] K. Panchanathan and R. Boyd. Indirect reciprocity can stabilize cooperation without the second-order free-rider problem. *Nature*, 432:499–502, 2004.

[16] M. Perc. Sustainable institutionalized punishment requires elimination of second-order free riders. *Scientific Reports*, 2:344, 2012.

[17] C. Hilbe and A. Traulsen. Emergence of responsible sanctions without second order free riders, antisocial punishment or spite. *Scientific Reports*, 2:458, 2012.

[18] 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.

[19] Saptarshi Pal and Christian Hilbe. Reputation effects drive the joint evolution of cooperation and social rewarding. *Nature Communications*, 13(1):5928, 2022.

[20] F. C. Santos and J. M. Pacheco. Risk of collective failure provides an escape from the tragedy of the commons. *Proceedings of the National Academy of Sciences USA*, 108:10421–10425, 2011.

[21] C. S. Gokhale and A. Traulsen. Strategy abundance in evolutionary many-player games with multiple strategies. *Journal of Theoretical Biology*, 238:180–191, 2011.

[22] J. Peña. Group size diversity in public goods games. *Evolution*, 66:623–636, 2012.

[23] 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.

[24] Peter D. Taylor. Evolutionarily stable strategies with two types of player. *Journal of Applied Probability*, 16(1):76–83, 1979.

[25] P. Schuster and K. Sigmund. Coyness, philandering and stable strategies. *Animal Behaviour*, 29:186–192, 1981.

[26] A. Gaunersdorfer, J. Hofbauer, and K. Sigmund. The dynamics of asymmetric games. *Theoretical Population Biology*, 29:345–357, 1991.

[27] J. Hofbauer. Evolutionary dynamics for bimatrix games: A Hamiltonian system? *Journal of Mathematical Biology*, 34:675–688, 1996.

[28] Josef Hofbauer and Ed Hopkins. Learning in perturbed asymmetric games. *Games and Economic Behavior*, 52(1):133–152, 2005.

[29] H. Ohtsuki. Stochastic evolutionary dynamics of bimatrix games. *Journal of Theoretical Biology*, 264:136–142, 2010.

[30] Alex McAvoy and Christoph Hauert. Asymmetric Evolutionary Games. *PLoS Computational Biology*, 11(8):1–26, 2015.

[31] C. Veller and L. K. Hayward. Finite-population evolution with rare mutations in asymmetric games. *Journal of Economic Theory*, 162:93–113, 2016.

[32] O.P. Hauser, C. Hilbe, K. Chatterjee, and M.A. Nowak. Social dilemmas among unequals. *Nature*, 572:524—527, 2019.

[33] M. Milinski, T. Röhl, and J. Marotzke. Cooperative interaction of rich and poor can be catalyzed by intermediate climate targets. *Climatic Change*, 109:807–814, 2011.

[34] Vitor V. Vasconcelos, Francisco C. Santos, Jorge M. Pacheco, and Simon A Levin. Climate policies under wealth inequality. *Proceedings of the National Academy of Sciences USA*, 111, 2014.

[35] M. Abou Chakra and A. Traulsen. Under high stakes and uncertainty the rich should lend the poor a helping hand. *Journal of Theoretical Biology*, 341:123–130, 2014.

[36] Ramona Merhej, Fernando P. Santos, Francisco S. Melo, and Francisco C. Santos. Cooperation and Learning Dynamics under Wealth Inequality and Diversity in Individual Risk Perception. *Journal of Artificial Intelligence Research*, 74:733–764, 2022.

[37] X Wang, M C Couto, N Wang, X An, B Chen, Y Dong, C Hilbe, and B Zhang. Cooperation and coordination in heterogeneous populations. *Philosophical Transactions of the Royal Society B*, 378, 2023.

[38] J. Maynard Smith and G. R. Price. The logic of animal conflict. *Nature*, 246:15–18, 1973.

[39] J. Maynard Smith. *Evolution and the Theory of Games*. Cambridge University Press, Cambridge, 1982.

[40] M. A. Nowak. *Evolutionary dynamics*. Harvard University Press, Cambridge MA, 2006.

[41] Tuomas W Sandholm and Robert H Crites. Multiagent reinforcement learning in the iterated prisoner's dilemma. *BioScience*, 37:147–166, Jul 1996.

[42] D. Fudenberg and D. Levine. *The Theory of Learning in Games*. MIT Press, Cambridge MA, 1998.

[43] M. W. Macy and A. Flache. Learning dynamics in social dilemmas. *Proceedings of the National Academy of Sciences USA*, 99:7229–7236, 2002.

[44] 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.

[45] M. Broom, C. Cannings, and G.T. Vickers. Multi-player matrix games. *Bulletin of Mathematical Biology*, 59(5):931–952, 1997.

[46] 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.

[47] P. D. Taylor and L. Jonker. Evolutionarily stable strategies and game dynamics. *Mathematical Biosciences*, 40:145–156, 1978.

[48] 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.

[49] C. S. Gokhale and A. Traulsen. Evolutionary games in the multiverse. *Proceedings of the National Academy of Sciences USA*, 107:5500–5504, 2010.

[50] Ross Cressman and Yi Tao. The replicator equation and other game dynamics. *Proceedings of the National Academy of Sciences USA*, 111:10810–10817, 2014.

[51] J. Peña, L. Lehmann, and G. Nöldeke. Gains from switching and evolutionary stability in multiplayer matrix games. *Journal of Theoretical Biology*, 346:23–33, 2014.

[52] 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.

[53] C. S. Gokhale and A. Traulsen. Mutualism and evolutionary multiplayer games: revisiting the Red King. *Proceedings of the Royal Society B*, 279:4611–4616, 2012.

[54] 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.

[55] Xinyu Zhang, Peng Peng, Yushan Zhou, Haifeng Wang, and Wenxin Li. Evolutionary Game-Theoretical Analysis for General Multiplayer Asymmetric Games. 2022.

[56] M. A. Nowak, A. Sasaki, C. Taylor, and D. Fudenberg. Emergence of cooperation and evolutionary stability in finite populations. *Nature*, 428:646–650, 2004.

[57] 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.

[58] D. Fudenberg, M. A. Nowak, C. Taylor, and L. A. Imhof. Evolutionary game dynamics in finite populations with strong selection and weak mutation. *Theoretical Population Biology*, 70:352–363, 2006.

[59] T. Sekiguchi and H. Ohtsuki. Fixation probabilities of strategies for bimatrix games in finite populations. *Dynamic Games and Applications*, 7:93–111, 2017.

[60] Takuya Sekiguchi. Fixation Probabilities of Strategies for Trimatrix Games and Their Applications to Triadic Conflict. *Dynamic Games and Applications*, 2022.

[61] S. Kurokawa and Y. Ihara. Emergence of cooperation in public goods games. *Proceedings of the Royal Society B*, 276:1379–1384, 2009.

[62] 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.

[63] 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.

[64] Takuya Sekiguchi. General conditions for strategy abundance through a self-referential mechanism under weak selection. *Physica A: Statistical Mechanics and its Applications*, 392(13):2886–2892, 2013.

[65] Bin Wu, Arne Traulsen, and Chaitanya S. Gokhale. Dynamic properties of evolutionary multiplayer games in finite populations. *Games*, 4(2):182–199, 2013.

[66] Dhaker Kroumi and Sabin Lessard. Average abundancy of cooperation in multi-player games with random payoffs. *Journal of Mathematical Biology*, 85(3):1–31, 2022.

[67] 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.

[68] 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.

[69] W. Barfuss, J. F. Donges, and J. Kurths. Deterministic limit of temporal difference reinforcement learning for stochastic games. *Physical Review E*, 99:043305, 2019.

[70] 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.

[71] 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.

[72] Alex McAvoy, Julian Kates-Harbeck, Krishnendu Chatterjee, and Christian Hilbe. Evolutionary instability of selfish learning in repeated games. *PNAS Nexus*, (July):1–15, 2022.

[73] Laura Schmid, Christian Hilbe, Krishnendu Chatterjee, and Nowak Martin A. Direct reciprocity between individuals that use different strategy spaces. *PLoS Computational Biology*, 18(6):1–29, 2022.

[74] Andrea Gaunersdorfer and Josef Hofbauer. Fictitious play, shapley polygons, and the replicator equation. *Games and Economic Behavior*, 11:279–303, Sep 1995.

[75] D. Fudenberg and L. A. Imhof. Imitation processes with small mutations. *Journal of Economic Theory*, 131:251–262, 2006.

[76] G. Wild and A. Traulsen. The different limits of weak selection and the evolutionary dynamics of finite populations. *Journal of Theoretical Biology*, 247:382–390, 2007.

[77] Joseph W. Baron and Tobias Galla. How successful are mutants in multiplayer games with fluctuating environments? Sojourn times, fixation and optimal switching. *Royal Society Open Science*, 5(3), 2018.

[78] Chai Molina and David J.D. Earn. Evolutionary stability in continuous nonlinear public goods games. *Journal of Mathematical Biology*, 74(1-2):499–529, 2017.

[79] Mark Broom and Jan Rychtář. A general framework for analysing multiplayer games in networks using territorial interactions as a case study. *Journal of Theoretical Biology*, 302:70–80, 2012.

[80] M. Perc, J. Gómez-Gardeñes, A. Szolnoki, L. M. Floría, and Y. Moreno. Evolutionary dynamics of group interactions on structured populations: A review. *Journal of The Royal Society Interface*, 10(80):20120997, 2013.

[81] Jorge Peña, Georg Nöldeke, and Laurent Lehmann. Evolutionary dynamics of collective action in spatially structured populations. *Journal of Theoretical Biology*, 382:122–136, 2015.

[82] J. Peña, B. Wu, and A. Traulsen. Ordering structured populations in multiplayer cooperation games. *Journal of The Royal Society Interface*, 13(114):20150881, 2016.

[83] Karan Pattni, Mark Broom, and Jan Rychtář. Evolutionary dynamics and the evolution of multiplayer cooperation in a subdivided population. *Journal of Theoretical Biology*, 429:105–115, 2017.

[84] Qi Su, Alex McAvoy, and Joshua B. Plotkin. Evolution of cooperation with contextualized behavior. *Science Advances*, 8(6):1–11, 2022.