



PAPER • OPEN ACCESS

Introspection dynamics: a simple model of counterfactual learning in asymmetric games

To cite this article: M C Couto *et al* 2022 *New J. Phys.* **24** 063010

View the [article online](#) for updates and enhancements.

You may also like

- [Dangerous drivers foster social dilemma structures hidden behind a traffic flow with lane changes](#)

Jun Tanimoto, Takuya Fujiki, Zhen Wang et al.

- [Reward versus punishment: averting the tragedy of the commons in eco-evolutionary dynamics](#)

Samrat Sohel Mondal, Mayank Pathak and Sagar Chakraborty

- [Coevolutionary dynamics of aspiration and strategy in spatial repeated public goods games](#)

Te Wu, Feng Fu and Long Wang



PAPER

Introspection dynamics: a simple model of counterfactual learning in asymmetric games

OPEN ACCESS

RECEIVED
13 January 2022REVISED
22 April 2022ACCEPTED FOR PUBLICATION
13 May 2022PUBLISHED
14 June 2022

Original content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the
title of the work, journal
citation and DOI.

M C Couto^{1,*} , S Giaimo²  and C Hilbe¹ ¹ Max Planck Research Group: Dynamics of Social Behavior, Max Planck Institute for Evolutionary Biology, 24306 Plön, Germany² Department of Evolutionary Theory, Max Planck Institute for Evolutionary Biology, 24306 Plön, Germany

* Author to whom any correspondence should be addressed.

E-mail: mccouto@evolbio.mpg.de**Keywords:** evolutionary game theory, counterfactual learning, myopic updating, asymmetric games, social dilemmas, volunteer's dilemma

Abstract

Social behavior in human and animal populations can be studied as an evolutionary process. Individuals often make decisions between different strategies, and those strategies that yield a fitness advantage tend to spread. Traditionally, much work in evolutionary game theory considers symmetric games: individuals are assumed to have access to the same set of strategies, and they experience the same payoff consequences. As a result, they can learn more profitable strategies by imitation. However, interactions are oftentimes asymmetric. In that case, imitation may be infeasible (because individuals differ in the strategies they are able to use), or it may be undesirable (because individuals differ in their incentives to use a strategy). Here, we consider an alternative learning process which applies to arbitrary asymmetric games, *introspection dynamics*. According to this dynamics, individuals regularly compare their present strategy to a randomly chosen alternative strategy. If the alternative strategy yields a payoff advantage, it is more likely adopted. In this work, we formalize introspection dynamics for pairwise games. We derive simple and explicit formulas for the abundance of each strategy over time and apply these results to several well-known social dilemmas. In particular, for the volunteer's timing dilemma, we show that the player with the lowest cooperation cost learns to cooperate without delay.

1. Introduction

Social behaviors often follow an evolutionary process. Behaviors that yield a high payoff proliferate, whereas inferior strategies go extinct [1, 2]. Providing a quantitative description of how individuals learn, make choices, and interact, however, is non-trivial. Actions of one group member may affect the behaviors of others; and, in a complex chain of interdependencies, this may, in turn, affect the group's surrounding environment [3–5]. Researchers aim to capture this complex dynamics with models of learning [6–13] and the tools of evolutionary game theory [14–16]. One key application of this theory is the study of cooperation in social dilemmas, where an individual's self-interest is at odds with the common good [17, 18]. For example, to be physically isolated during a pandemic is an action that is simultaneously costly for the individual and valuable for public health. Game theory is a standard mathematical framework to model decision-making in this context [19]. A *game*—defined by the players, their possible strategies, and their payoffs—captures how individuals interact and what the resulting consequences are.

One of the more subtle aspects of such a model is how it specifies the strategy updating process [20, 21]. In evolutionary game theory, individuals do not act optimally from the outset. Rather they dynamically adjust their strategies, based on their current payoffs. To capture this dynamics, the model needs to specify how individuals adopt new strategies, which is governed by the players' *strategy update rule* or *selection rule*. One example of such an update rule is imitation by pairwise comparison [22–24]: occasionally, individuals compare their own payoff to the payoff of a co-player. The larger the co-player's payoff, the more likely the focal individual imitates the co-player's strategy. Alternatively, there is also work that considers

conformity-driven imitation [25–27]. Here, individuals tend to copy those strategies that are abundant among their neighbors, independent of the strategies' payoff consequences.

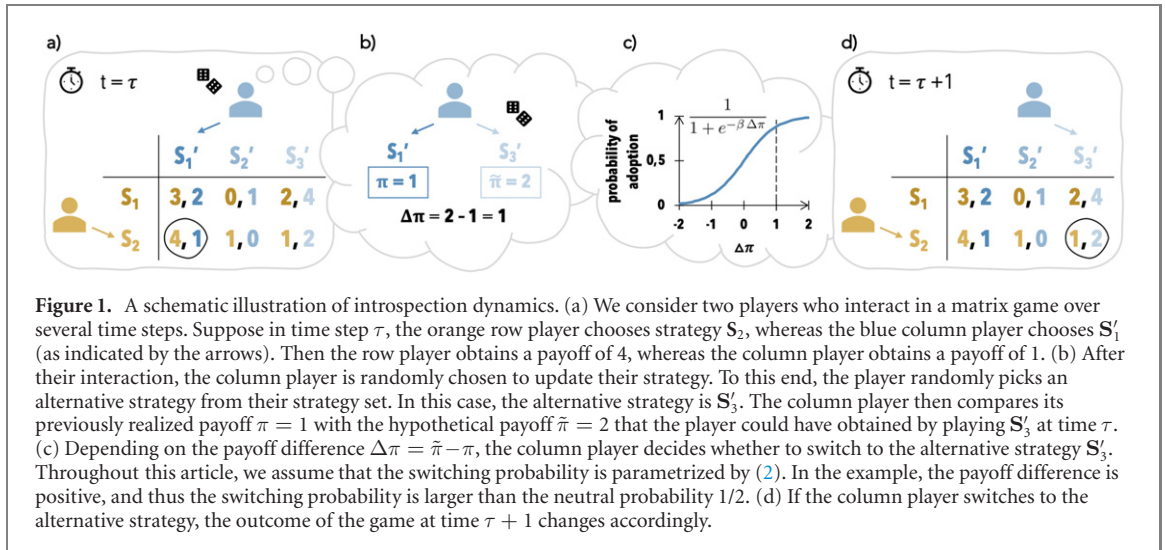
Yet 'not all learning is social learning' (i.e., learning from others) [2]. People can also adjust their strategies via introvert reasoning. For example, some models suggest that individuals may experiment with new strategies when their current payoff falls below a predefined aspiration level [28–30]. Other models assume that individuals adopt new strategies by figuring out the optimal response to their co-players' strategies, as in fictitious play or different variants of best reply dynamics [31–33]. Additionally, there is a body of work that studies the interplay and competition among different strategy updating rules [34–36].

Models of introvert reasoning are particularly relevant in asymmetric games [37, 38], where individuals are inherently unequal. Such inequalities naturally arise when individuals take different roles in an interaction (as in the ultimatum game [39, 40]), or when they differ in their available strategies [41, 42]. But even if individuals have access to the same set of strategies, they may differ in other relevant aspects, such as their initial endowments [43, 44], costs [45], productivities [46], or their location in a social network or hierarchy [47–49]. In the presence of such heterogeneities, strategies that yield a high payoff for one player may no longer be optimal for another. In this case, at least some degree of introvert reasoning seems to be required for effective learning.

In the following, we consider a learning model based on introvert reasoning that applies to all pairwise games. We assume that, from time to time, one individual is given the opportunity to change their strategy. In doing so, the individual thinks counterfactually: 'how much better off a random alternative strategy would have made me in the last interaction?' The larger the difference between the counterfactual payoff and the realized payoff, the more likely the individual is to abandon the current strategy in favor of the random alternative. We refer to this strategy update rule as *introspection dynamics*. Similar to previous studies [50–53], introspection dynamics assumes that players are 'myopic'. That is, when considering switching to a different strategy, players assume their co-player's strategy to be fixed. In particular, they do not anticipate that their new strategy may motivate their co-player to adapt later on. The rule implemented by introspection dynamics was previously used to simulate learning processes in (repeated) social dilemmas [52–55]. Here, we introduce an easy-to-implement model in the context of one-shot games, and we systematically analyze its mathematical properties. In particular, we derive an explicit formula for how likely players are to use any given strategy in the long run. This formula takes a particularly simple form when selection is weak (i.e., when realized payoffs only marginally affect a player's learning behavior).

Although symmetric games are more commonly studied, researchers have been interested in the evolutionary dynamics of asymmetric games for decades [56–66]. In comparison to these previous models, introspection dynamics has several properties that make it particularly comfortable to work with. For example, in models based on best reply dynamics and fictitious play [7, 31], players need to be aware of the entire payoff matrix. Moreover, they need to be able to compute best responses at all updating steps. Instead, in introspection dynamics, individuals only compare two given strategies at any point in time. In this way, introspection dynamics makes weaker cognitive assumptions on the players, and it allows a more immediate characterization of the eventual strategy distribution. This advantage seems to be particularly valuable in games with many strategies [67]. Similarly, there are natural connections between introspection dynamics and classical evolutionary models of asymmetric games [56–64]. For instance, a common approach is to assume that each participant in an asymmetric game is a member of a distinct subpopulation. Evolution then only occurs within each subpopulation. This modeling approach, however, is difficult to apply when all players are unique (for example, when all players in a population differ in how much they benefit from a public good [54]), or when there are no clear boundaries between different subpopulations. In such cases, introspection dynamics may provide a more natural way to explore the resulting strategy dynamics.

The remainder of this paper is organized as follows. In section 2, we formalize introspection dynamics for 2-player, 1-shot games with finitely many strategies. Section 3 contains our main analytical results. Here, we compute the unique stationary distribution of strategies. In addition, we provide explicit formulas for two cases: when selection is weak, and when players can choose among two strategies only. To illustrate these results, section 4 discusses the effect of asymmetry in several well-known social dilemmas. In particular, for the volunteer's timing dilemma, we show that the player with the lower cooperation cost tends to cooperate without delay. We also look at symmetric games and compare introspection dynamics with an evolutionary model of imitation [24]. Finally, section 5 discusses the results and puts them into a broader context. All proofs are provided in the appendices.



2. Model

To model asymmetric interactions, we consider two individuals, player 1 and player 2, who interact in a normal-form game. Player 1 can choose among m pure strategies, S_1, \dots, S_m , and player 2 can choose among n strategies, S'_1, \dots, S'_n . If player 1 chooses strategy S_i and player 2 chooses strategy S'_j , the resulting payoffs are π_{ij} and π'_{ij} , respectively. We represent such a game by a bi-matrix,

$$\begin{array}{c|ccc}
 & S'_1 & \dots & S'_n \\
 \hline
 S_1 & \pi_{11}, \pi'_{11} & \dots & \pi_{1n}, \pi'_{1n} \\
 \vdots & \vdots & \ddots & \vdots \\
 S_m & \pi_{m1}, \pi'_{m1} & \dots & \pi_{mn}, \pi'_{mn}
 \end{array} \tag{1}$$

In this representation, player 1 can choose one of the m rows, whereas player 2 chooses one of the n columns. Thus, we can also refer to player 1 as the ‘row player’, and to player 2 as the ‘column player’. The first entry in the bi-matrix is the payoff to player 1, whereas the second entry is the payoff to player 2. The game is called symmetric if the two players have the same set of strategies (that is, if $m = n$ and $S_i = S'_i$ for all i), and if the corresponding strategies yield the same payoffs (if $\pi_{ij} = \pi'_{ji}$ for all i, j). In what follows we mainly focus on asymmetric games.

We assume that the two players interact over many time steps, engaging each time in the game defined by the above bi-matrix. In each time step, they can dynamically adjust their strategies. A possible learning dynamics may posit that players compare their current strategy with another one, assuming that the co-player’s strategy remains fixed. The process, which we refer to as *introspection dynamics*, goes as in figure 1. At each time step, one player is randomly chosen to reconsider their strategy. To this end, the player randomly draws an alternative strategy from the set of all its possible strategies. The player then compares its realized payoff π with the payoff $\tilde{\pi}$ the player could have obtained by playing the alternative strategy instead (keeping the co-player’s strategy fixed). Let $\Delta\pi := \tilde{\pi} - \pi$ be the difference between the counterfactual payoff and the realized payoff. The player switches to the alternative strategy with probability $\varphi_\beta(\Delta\pi)$ given by the Fermi function [23, 24, 68],

$$\varphi_\beta(\Delta\pi) = \frac{1}{1 + e^{-\beta \Delta\pi}}. \tag{2}$$

Here, $\beta \geq 0$ is a parameter that measures the intensity of selection or selection strength. In one limiting case, $\beta \rightarrow 0$, payoffs are irrelevant for the learning process and any alternative strategy is adopted with probability $1/2$. In the other limiting case $\beta \rightarrow \infty$, players only adopt the alternative strategy if the resulting payoff matches at least the current payoff. We refer to these two limits as the case of weak selection and strong selection, respectively.

As we iterate this elementary updating step over time, we obtain a stochastic process on the space of all strategy profiles $(S_1, S'_1), (S_1, S'_2), \dots, (S_m, S'_n)$. Each state corresponds to the strategies that the two players

currently adopt. Simulations of this process are straightforward [54, 55]. Here, our goal is to analyze the mathematical properties of this process.

3. Analytical properties of introspection dynamics

3.1. An explicit formula for the stationary strategy distribution

To derive explicit results, we note that each player’s updating behavior only depends on the players’ current strategies (whereas it is independent of the previous strategies of the players). As a result, we can represent the dynamics by a Markov chain. Given the current state $(\mathbf{S}_i, \mathbf{S}'_j)$, we can compute the probability that the state changes to $(\mathbf{S}_k, \mathbf{S}'_l)$ in one time step. This transition probability is

$$T_{ij,kl} = \begin{cases} \frac{1}{2(m-1)} \cdot \varphi_\beta(\pi_{kj} - \pi_{ij}) & \text{if } i \neq k \wedge j = l \text{ (player 1 changes)} \\ \frac{1}{2(n-1)} \cdot \varphi_\beta(\pi'_{il} - \pi'_{ij}) & \text{if } i = k \wedge j \neq l \text{ (player 2 changes)} \\ 0 & \text{if } i \neq k \wedge j \neq l \text{ (both players change)} \\ 1 - \frac{1}{2} \left(\sum_{p \neq i}^m \frac{\varphi_\beta(\pi_{pj} - \pi_{ij})}{m-1} + \sum_{q \neq j}^n \frac{\varphi_\beta(\pi'_{iq} - \pi'_{ij})}{n-1} \right) & \text{if } i = k \wedge j = l \text{ (no player changes).} \end{cases} \quad (3)$$

In these expressions, the factor $\frac{1}{2}$ corresponds to randomly drawing one of the two players. Similarly, the factors $1/(m-1)$ and $1/(n-1)$ correspond to randomly drawing an alternative strategy. The expressions $\varphi_\beta(\Delta\pi)$ then give the probability that the alternative strategy is adopted Eq. (2). We can collect these transition probabilities in an $mn \times mn$ matrix $T = (T_{ij,kl})$. Here, the first double index denotes the row that corresponds to the previous state $(\mathbf{S}_i, \mathbf{S}'_j)$ of the Markov chain, whereas the second double index corresponds to the next state $(\mathbf{S}_k, \mathbf{S}'_l)$. By construction, T is nonnegative and row stochastic (see appendix A).

We can use the matrix T to describe how likely we are to find the process in a given state at some time t given the initial distribution of the process. Let $v_{ij}(t)$ denote the probability that at time t , the process is in state $(\mathbf{S}_i, \mathbf{S}'_j)$. The mn -dimensional row-vector $\mathbf{v}(t)$ collects these probabilities in the same order of states as in the transition matrix. By the theory of Markov chains, the strategy distribution at time t is

$$\mathbf{v}(t) = \mathbf{v}(0) T^t, \quad (4)$$

where $\mathbf{v}(0)$ is the initial strategy distribution.

To describe the long-run dynamics, we assume in the following that the selection strength β is finite (even if it may be arbitrarily large). In that case, the transition matrix T is primitive (see proposition 1 in appendix A). Thus, it follows from the Perron–Frobenius theorem [69] that $\mathbf{v}(t)$ converges in time to a unique and positive stationary distribution \mathbf{u} , which is independent of the initial distribution $\mathbf{v}(0)$. The stationary distribution \mathbf{u} solves the eigenvector problem

$$\begin{aligned} \mathbf{u} &= \mathbf{u} T \\ \mathbf{u} \mathbf{e}^\top &= 1. \end{aligned} \quad (5)$$

Here, \mathbf{e} denotes the mn -dimensional row-vector where each entry is equal to 1 and the superscript \top indicates transposition. Hence, the second equation in (5) is the usual normalization for a probability vector (requiring that the sum of all components of \mathbf{u} is equal to 1). While (5) defines the stationary distribution implicitly, we can also obtain an explicit formula. To this end, we multiply the second equation in (5) by the row-vector \mathbf{e} from the right, and add the result to the first equation. After some rearranging, this yields the relationship $\mathbf{u} (I + U - T) = \mathbf{e}$. Here, $U = \mathbf{e}^\top \mathbf{e}$ is the $mn \times mn$ matrix with all entries being equal to 1, and I is the identity matrix. The matrix $(I + U - T)$ is invertible (see proposition 2 in appendix A, with $(I + U - T)^{-1}$ being a *fundamental* matrix of the ergodic chain [70]). Thus, we obtain the explicit representation

$$\mathbf{u} = \mathbf{e} (I + U - T)^{-1}. \quad (6)$$

This equation allows us to numerically compute the stationary distribution $\mathbf{u} = (u_{ij})$ of introspection dynamics for any finite normal-form game.

To derive expressions that more immediately relate \mathbf{u} to the payoffs of the game, we consider two important special cases in the following. First, for any number of strategies, we derive an approximation in

the limit of weak selection ($\beta \ll 1$). After that, we provide exact formulas in the case where each player can choose among two strategies ($m = n = 2$).

3.2. Weak selection

To directly relate the long-run abundance of strategies with the game’s payoffs, we consider the limit of weak selection. For this, we first note that both the transition matrix (3) and its stationary distribution (6) depend on β ; to make this dependence more explicit, we write $T(\beta)$ for the transition matrix and $\mathbf{u}(\beta)$ for the stationary distribution. We then expand $\mathbf{u}(\beta)$ and $T(\beta)$ as a Taylor series around $\beta = 0$,

$$\begin{aligned} T(\beta) &= T_0 + T_1\beta + \mathcal{O}(\beta^2) \\ \mathbf{u}(\beta) &= \mathbf{u}_0 + \mathbf{u}_1\beta + \mathcal{O}(\beta^2). \end{aligned} \tag{7}$$

Here, $T_0 := T|_{\beta=0}$ and $T_1 := \partial T/\partial\beta|_{\beta=0}$ are the constant and the linear term of the Taylor expansion of the transition matrix. Both terms can be computed explicitly, see appendix B. Similarly, $\mathbf{u}_0 := \mathbf{u}|_{\beta=0}$ and $\mathbf{u}_1 := \partial\mathbf{u}/\partial\beta|_{\beta=0}$ are the constant and the linear term of the stationary distribution; those are the terms we wish to compute. To this end, we take the expressions (7) and plug them into the eigenvector problem (5). This yields the relation

$$\begin{aligned} \mathbf{u}_0 + \mathbf{u}_1\beta + \mathcal{O}(\beta^2) &= \mathbf{u}_0T_0 + (\mathbf{u}_0T_1 + \mathbf{u}_1T_0)\beta + \mathcal{O}(\beta^2), \\ \mathbf{u}_0\mathbf{e}^\top + \mathbf{u}_1\mathbf{e}^\top\beta + \mathcal{O}(\beta^2) &= 1. \end{aligned} \tag{8}$$

By setting $\beta = 0$ in (8), we note that \mathbf{u}_0 needs to satisfy the linear system

$$\begin{aligned} \mathbf{u}_0 &= \mathbf{u}_0 T_0 \\ \mathbf{u}_0 \mathbf{e}^\top &= 1. \end{aligned} \tag{9}$$

Similarly, by taking the first derivative of both sides of (8) with respect to β , and then setting $\beta = 0$, it follows that \mathbf{u}_1 needs to satisfy

$$\begin{aligned} \mathbf{u}_1 (I - T_0) &= \mathbf{u}_0 T_1. \\ \mathbf{u}_1 \mathbf{e}^\top &= 0. \end{aligned} \tag{10}$$

As we show in proposition 3 in appendix B, both systems (9) and (10) can be solved explicitly.

In the special case that both players have the same number of strategies, $m = n$, the solution becomes particularly simple. In that case, we can approximate the abundance of strategy profile $(\mathbf{S}_i, \mathbf{S}'_j)$ in the stationary strategy distribution by

$$(\mathbf{u})_{ij} = \frac{1}{n^2} + \frac{1}{2n^4} \sum_{p=1}^n \sum_{q=1}^n ((\pi_{ij} + \pi_{iq} - \pi_{pj} - \pi_{pq}) + (\pi'_{ij} + \pi'_{pj} - \pi'_{iq} - \pi'_{pq})) \beta + \mathcal{O}(\beta^2). \tag{11}$$

To interpret this formula, let us introduce the following shortcut notation for some payoff averages for player 1,

$$\pi_{\bullet j} = \frac{1}{n} \sum_{i=1}^n \pi_{ij}, \quad \pi_{i\circ} = \frac{1}{n} \sum_{j=1}^n \pi_{ij}, \quad \pi_{\bullet\circ} = \frac{1}{n^2} \sum_{i,j=1}^n \pi_{ij}. \tag{12}$$

Here, $\pi_{\bullet j}$ is the average payoff player 1 obtains when randomly sampling a strategy against a co-player with strategy \mathbf{S}'_j . The next expression, $\pi_{i\circ}$ is player 1’s average payoff when using strategy \mathbf{S}_i against a randomly sampling co-player. Finally, $\pi_{\bullet\circ}$ is player 1’s average payoff if both players sample randomly. Analogous averages can be defined with respect to player 2’s payoffs π' . Using this notation, we can rewrite the weak-selection formula (11) as

$$(\mathbf{u})_{ij} \approx \frac{1}{n^2} + \frac{\beta}{2n^2} ((\pi_{ij} - \pi_{\bullet j}) + (\pi_{i\circ} - \pi_{\bullet\circ}) + (\pi'_{ij} - \pi'_{i\circ}) + (\pi'_{\bullet j} - \pi'_{\bullet\circ})). \tag{13}$$

We say the strategy profile $(\mathbf{S}_i, \mathbf{S}'_j)$ is favored by selection if its abundance is larger than neutral, $(\mathbf{u})_{ij} > 1/n^2$. For weak selection, expression (13) suggests the following two mechanisms for a strategy profile to be favored: either (i) player 1’s payoff from using strategy \mathbf{S}_i against \mathbf{S}'_j is better than average, $\pi_{ij} > \pi_{\bullet j}$; or (ii) player 1’s payoff from using strategy \mathbf{S}_i against a random strategy of the co-player is better than average, $\pi_{i\circ} > \pi_{\bullet\circ}$. Two analogous mechanisms apply to player 2.

Interestingly, similar results have been previously derived for a birth–death model [60]. In that model, members of two separate populations engage in a bi-matrix game. The members of the first population act

in the role of the row player, whereas the members of the second population act as the column player. In the special case that mutations are rare and selection is weak, the results of that model coincide with ours. In that case, the likelihood that population 1 uses strategy S_i and that population 2 uses strategy S'_j simplifies to our Eq. (13). This agreement can be regarded as another instance of a more general observation: in the limit of weak selection, different selection rules often turn out to be equivalent [71, 72].

The previous expressions (11) and (13) tell us how often we are to observe a strategy profile (S_i, S'_j) on average. In many applications, it is also relevant to know how often player 1 adopts the given strategy S_i , irrespective of the co-player's strategy. Based on (11), we can approximate the corresponding marginal probability $\xi_i := \sum_{j=1}^n u_{ij}$ as

$$\xi_i = \frac{1}{n} + \frac{\beta}{n^3} \sum_{p=1}^n \sum_{q=1}^n (\pi_{iq} - \pi_{pq}) + \mathcal{O}(\beta^2). \tag{14}$$

Using the notation (12), this expression simplifies to $\xi_i = 1/n + \beta(\pi_{i\bullet} - \pi_{\bullet\bullet})/n + \mathcal{O}(\beta^2)$. In particular, we can use this formula to predict which of two strategies S_i and S_k is more likely to be played over time. We obtain that $\xi_i > \xi_k$ if and only if $\pi_{i\bullet} > \pi_{k\bullet}$ (i.e., if and only if S_i performs better than S_k against a uniform sample of the co-player's strategies). Interestingly, this condition naturally depends on player 1's own payoffs π_{pq} , but it is independent of the co-player's payoffs π'_{pq} . While such a result may appear intuitive, we show that it only holds under weak selection. Once selection becomes stronger, the co-player's payoffs can have a major impact on how likely the focal player is to play a certain strategy (for an explicit example, see appendix B).

In this section, we have derived the first-order approximation of the stationary distribution as a function of the selection strength β . Using similar methods, we can recursively compute all higher-order terms in the Taylor expansion of \mathbf{u} . The respective expressions are derived in appendix C.

3.3. Games with two strategies

We now apply our model of introspection dynamics to general asymmetric 2-player, 2-strategy games. In contrast to the previous section, the results here are valid for any intensity of selection. For $m = n = 2$, the payoff matrix (1) simplifies to

| | | | |
|-------|-----------------------|-----------------------|--|
| | S'_1 | S'_2 | |
| S_1 | π_{11}, π'_{11} | π_{12}, π'_{12} | |
| S_2 | π_{21}, π'_{21} | π_{22}, π'_{22} | |

(15)

Since the transitions in introspection dynamics only depend on payoff differences between strategies, we can further simplify this payoff matrix. Appendix A shows that the transition probabilities in (3) remain unchanged if we add a constant to all payoffs π_{ij} in a given column j or to all payoffs π'_{ij} in a given row i . We can therefore assume without loss of generality that the payoff matrix (15) takes the form

| | | | |
|-------|---------|---------|--|
| | S'_1 | S'_2 | |
| S_1 | A, A' | $0, 0$ | |
| S_2 | $0, 0$ | B, B' | |

(16)

where $A := \pi_{11} - \pi_{21}$, $A' := \pi'_{11} - \pi'_{12}$, $B := \pi_{22} - \pi_{12}$, and $B' := \pi'_{22} - \pi'_{21}$. Thus, the number of free parameters reduces from eight in (15) to four in (16). For general $m \times n$ payoff matrices, this approach reduces the number of free parameters from $2mn$ to $2mn - m - n$.

Using the payoff matrix (16), we can write the transition matrix T according to (3) explicitly,

$$\begin{pmatrix} 1 - \frac{1}{2(1 + e^{\beta A})} - \frac{1}{2(1 + e^{\beta A'})} & \frac{1}{2(1 + e^{\beta A'})} & \frac{1}{2(1 + e^{\beta A})} & 0 \\ \frac{1}{2(1 + e^{-\beta A'})} & 1 - \frac{1}{2(1 + e^{-\beta A'})} - \frac{1}{2(1 + e^{-\beta B})} & 0 & \frac{1}{2(1 + e^{-\beta B})} \\ \frac{1}{2(1 + e^{-\beta A})} & 0 & 1 - \frac{1}{2(1 + e^{-\beta A})} - \frac{1}{2(1 + e^{-\beta B'})} & \frac{1}{2(1 + e^{-\beta B'})} \\ 0 & \frac{1}{2(1 + e^{\beta B})} & \frac{1}{2(1 + e^{\beta B'})} & 1 - \frac{1}{2(1 + e^{\beta B})} - \frac{1}{2(1 + e^{\beta B'})} \end{pmatrix}.$$

We then obtain the stationary strategy distribution $\mathbf{u} = (u_{11}, u_{12}, u_{21}, u_{22})$ by either solving (5) or using (6),

$$\begin{aligned} u_{11} &= \frac{1}{C} \left(e^{(A+B)\beta} + e^{(A'+B')\beta} + 2e^{(A+A')\beta} + 2e^{(A+A'+B)\beta} + 2e^{(A+A'+B')\beta} \right) \\ u_{12} &= \frac{1}{C} \left(2e^{A\beta} + e^{(A+B)\beta} + 2e^{B'\beta} + e^{(A'+B')\beta} + 2e^{(A+B')\beta} \right) \\ u_{21} &= \frac{1}{C} \left(2e^{B\beta} + e^{(A+B)\beta} + 2e^{A'\beta} + e^{(A'+B')\beta} + 2e^{(A'+B)\beta} \right) \\ u_{22} &= \frac{1}{C} \left(e^{(A+B)\beta} + e^{(A'+B')\beta} + 2e^{(B+B')\beta} + 2e^{(A+B+B')\beta} + 2e^{(A'+B+B')\beta} \right). \end{aligned} \tag{17}$$

Here, C is a normalization constant ensuring that components of \mathbf{u} add up to 1. Accordingly, the respective marginal probabilities $\xi_1 = u_{11} + u_{12}$ and $\xi'_1 = u_{11} + u_{21}$ are

$$\begin{aligned} \xi_1 &= \frac{2}{C} \left(1 + e^{A'\beta} \right) \left(e^{A\beta} + e^{B'\beta} + e^{(A+B)\beta} + e^{(A+B')\beta} \right) \\ \xi'_1 &= \frac{2}{C} \left(1 + e^{A\beta} \right) \left(e^{A'\beta} + e^{B\beta} + e^{(A'+B')\beta} + e^{(A'+B)\beta} \right). \end{aligned}$$

The remaining marginal probabilities are $\xi_2 = 1 - \xi_1$ and $\xi'_2 = 1 - \xi'_1$. We note that, in general, the stationary distribution \mathbf{u} does not factorize over these marginal distributions. That is, in general, $u_{ij} \neq \xi_i \cdot \xi'_j$ for $i, j \in \{1, 2\}$ and $\beta > 0$. However, one can verify that such a factorization holds when the game's payoffs satisfy $A = -B$ and $A' = -B'$. These latter two conditions imply that each player's incentive to choose one strategy (rather than the other) is independent of the opponent's strategy. In such a case, this independence is also reflected in the game's stationary distribution.

The previous results hold for arbitrary selection strength. We can also look at the weak selection approximation. Using (13), we obtain

$$\begin{aligned} u_{11} &\approx \frac{1}{4} + \frac{\beta}{32}(3A + 3A' - B - B') \\ u_{12} &\approx \frac{1}{4} + \frac{\beta}{32}(A - 3A' - 3B + B') \\ u_{21} &\approx \frac{1}{4} + \frac{\beta}{32}(-3A + A' + B - 3B') \\ u_{22} &\approx \frac{1}{4} + \frac{\beta}{32}(-A - A' + 3B + 3B'). \end{aligned} \tag{18}$$

In this case, the resulting marginal probabilities become particularly simple, $\xi_1 = 1/2 + \beta(A - B)/8$ and $\xi'_1 = 1/2 + \beta(A' - B')/8$. Player 1 favors strategy \mathbf{S}_1 if and only if $A > B$, and the analogous result holds for player 2. In the following, we apply these formulas to discuss the introspection dynamics of some classical 2-strategy games.

4. Applications

4.1. Social dilemmas with two strategies

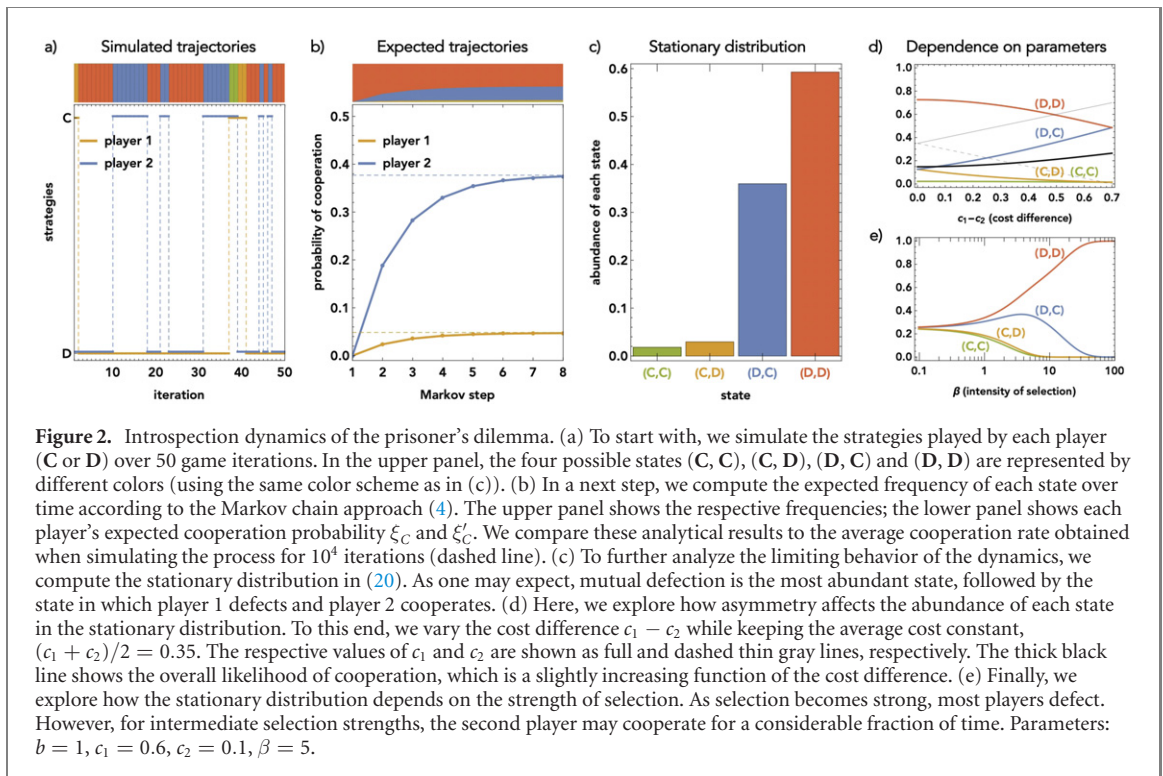
As a first application of our model, we explore the introspection dynamics of asymmetric social dilemmas. In social dilemmas, players can choose whether to cooperate (C) or to defect (D). When players choose the same action, they prefer mutual cooperation to mutual defection; yet when they choose different actions, a defector gets a higher payoff than a cooperator [17].

4.1.1. Prisoner's dilemma

To start with, we consider the most stringent form of a social dilemma, the prisoner's dilemma [18]. For a simple instantiation of this game, we assume that a cooperating player pays a cost in order for the co-player to get a benefit. We incorporate asymmetry by assuming that the cooperation costs $c_i > 0$ may differ between players (whereas the benefit b of cooperation is the same for both). Therefore, the payoff matrix takes the form

| | | |
|---|--------------------|-----------|
| | C | D |
| C | $b - c_1, b - c_2$ | $-c_1, b$ |
| D | $b, -c_2$ | $0, 0$ |

(19)



The unique Nash equilibrium of this game is for both players to defect, independent of the players' exact costs. For an easier interpretation, however, we assume without loss of generality that cooperation tends to be more costly for the first player, $c_1 \geq c_2$.

In the following, we ask how likely players are to cooperate when they update their strategies according to introspection dynamics. To this end, we first consider the case of a fixed benefit of $b = 1$. Moreover, we assume that the first player faces considerable cooperation costs, $c_1 = 0.6$, whereas the second player's costs are negligible, $c_2 = 0.1$. To illustrate the workings of introspection dynamics, we start by simulating the basic process described in section 2. Figure 2(a) shows a representative realization. Over time, the two players independently switch between cooperation and defection. As a result, they experience all possible outcomes: there are times in which both players defect, but also instances in which either one or both players cooperate. Overall, however, mutual defection appears to be most abundant, as one may expect.

To obtain a more quantitative understanding, we compute how likely we are to observe each of the four possible outcomes over time. To this end, we assume that initially both players defect, such that $\mathbf{v}(0) = (v_{CC}(0), v_{CD}(0), v_{DC}(0), v_{DD}(0)) = (0, 0, 0, 1)$. Then we use (4) to compute $\mathbf{v}(t)$ for all future time steps. In figure 2(b), we show the resulting cooperation probability for each player, as defined by $\xi_C(t) = v_{CC}(t) + v_{CD}(t)$ for the first player and $\xi'_C(t) = v_{CC}(t) + v_{DC}(t)$ for the second player. As before, we observe that both players are most likely to defect. However, while player 1's cooperation probability remains low (approximately 5%), player 2's cooperation probability quickly reaches a stable value of about 38%.

To further explore this limiting behavior, we compute the game's stationary distribution. Setting the four game parameters in the payoff matrix (16) as $A = -c_1$, $A' = -c_2$, $B = c_1$, $B' = c_2$, we can use (17) to compute the stationary distribution $\mathbf{u} = (u_{CC}, u_{CD}, u_{DC}, u_{DD})$ as

$$\mathbf{u} = \frac{1}{(1+e^{\beta c_1})(1+e^{\beta c_2})} \cdot (1, e^{\beta c_2}, e^{\beta c_1}, e^{\beta(c_1+c_2)}). \quad (20)$$

In particular, we note that the invariant distribution does not depend on the benefit of cooperation (since b does not enter the payoff differences A, A', B, B'). Moreover, because $\beta \geq 0$ and $c_1 \geq c_2 > 0$, the abundances of the four possible states always obey the same relationship as in figure 2(c),

$$u_{CC} \leq u_{CD} \leq u_{DC} \leq u_{DD}. \quad (21)$$

For $\beta > 0$ and $c_1 > c_2$ all inequalities in (21) are strict. That is, among the two players, the player with the lower cooperation cost can be expected to cooperate more often. However, mutual defection is always the most abundant state.

We can also use the stationary distribution (20) to study how parameter changes affect cooperation. In figure 2(d), we explore the effect of asymmetry. To this end, we increase the difference in the players' costs $c_1 - c_2$ while keeping the average cost $(c_1 + c_2)/2$ fixed. Interestingly, we observe a weakly positive effect. As we increase c_1 and decrease c_2 , the first player's reduced cooperation is more than compensated by the second player.

Similarly, in figure 2(e) we explore the impact of selection strength. When selection is weak, such that payoff differences have a negligible impact, the stationary distribution simplifies to

$$\mathbf{u} \approx \left(\frac{1}{4} - \frac{c_1+c_2}{8} \cdot \beta, \frac{1}{4} - \frac{c_1-c_2}{8} \cdot \beta, \frac{1}{4} + \frac{c_1-c_2}{8} \cdot \beta, \frac{1}{4} + \frac{c_1+c_2}{8} \cdot \beta \right). \tag{22}$$

In particular, the last two states (D, C) and (D, D) are favored by selection, whereas the other two states are disfavored.

In contrast, for the strong selection regime, we take the limit of \mathbf{u} as β becomes arbitrarily large. We obtain

$$\lim_{\beta \rightarrow \infty} \mathbf{u} = (0, 0, 0, 1). \tag{23}$$

Thus, we recover the classical prediction that both players learn to defect at all times.

4.1.2. Stag-hunt game

As another instance of a social dilemma, we explore a version of the stag-hunt game [73]. In contrast to the prisoner's dilemma, we now assume that players only derive a benefit if they both cooperate (which could reflect the advantage of collective action in hunting expeditions). Under this assumption, the payoff matrix becomes

| | | | |
|---|--------------------|-----------|--|
| | C | D | |
| C | $b - c_1, b - c_2$ | $-c_1, 0$ | |
| D | $0, -c_2$ | $0, 0$ | |

(24)

As before, we assume the payoff parameters satisfy $b > c_1 \geq c_2 > 0$. Under this assumption, the stag-hunt game becomes a coordination game. There are two pure equilibria, according to which either both players cooperate, or both players defect. Mutual cooperation is always payoff-dominant (it is the equilibrium that gives a higher payoff to both players). However, when the benefit of cooperation is comparably small, $b < c_1 + c_2$, mutual defection is risk-dominant (loosely meaning that mutual defection is the safer option when there is uncertainty about the co-player's actions [74]).

We can analyze the introspection dynamics of the stag-hunt game in the same way as the prisoner's dilemma. This time, we set $A = b - c_1, A' = b - c_2, B = c_1, B' = c_2$. As a result, the stationary distribution according to (17) becomes

$$\mathbf{u} = \frac{1}{e^{\beta b} + e^{\beta c_1} + e^{\beta c_2} + e^{\beta(c_1+c_2)}} \cdot (e^{\beta b}, e^{\beta c_2}, e^{\beta c_1}, e^{\beta(c_1+c_2)}). \tag{25}$$

It follows that there are two possible orderings of the abundance of the four states. Which ordering applies depends on how the benefit b relates to the total costs of cooperation $c_1 + c_2$,

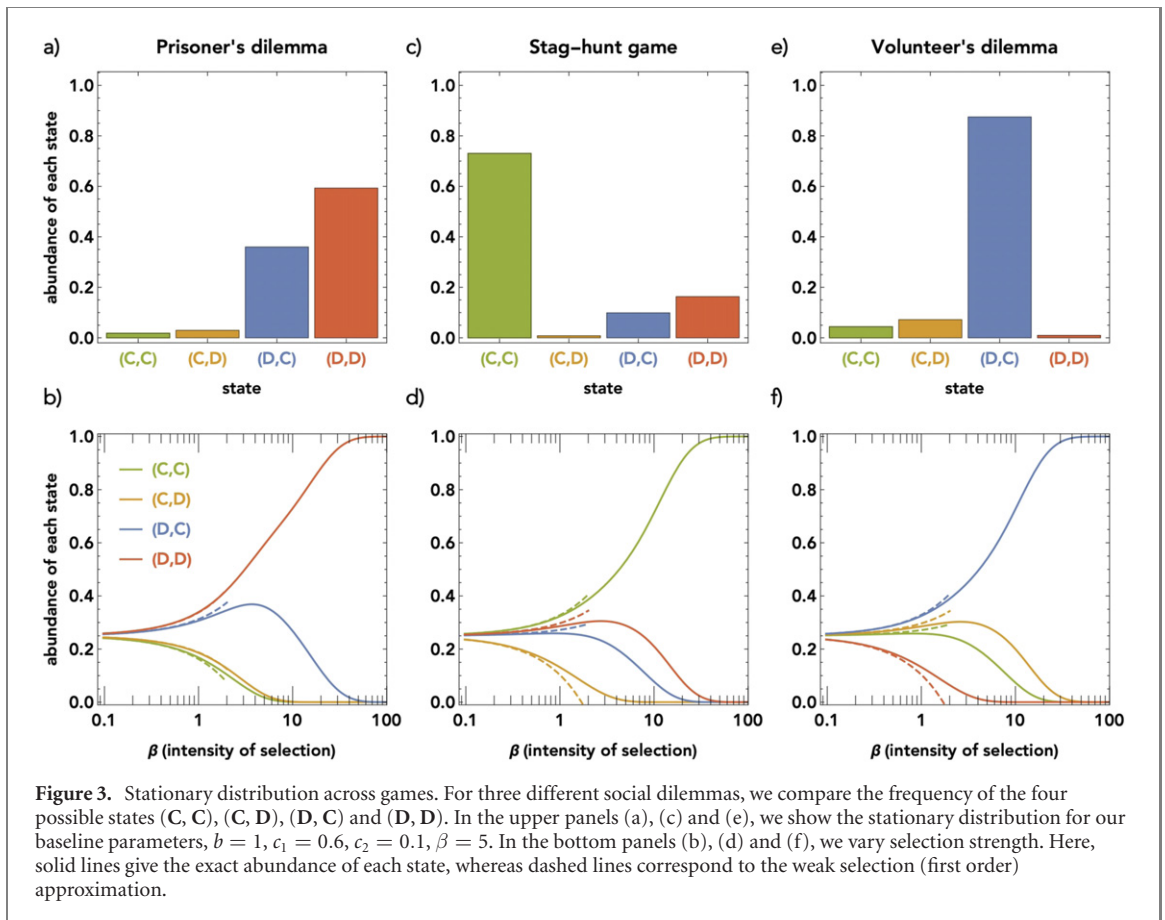
$$\begin{aligned} u_{CD} \leq u_{DC} \leq u_{DD} \leq u_{CC} & \quad \text{if } b \geq c_1 + c_2 \\ u_{CD} \leq u_{DC} \leq u_{CC} \leq u_{DD} & \quad \text{if } b \leq c_1 + c_2. \end{aligned}$$

Irrespective of the precise ordering, players are always most likely to settle at one of the two equilibrium outcomes. Moreover, introspection dynamics lends further support to the static notion of risk-dominance: players only coordinate on mutual cooperation if the benefit exceeds the sum of the costs. In figure 3, we illustrate this result for our baseline parameters, $b = 1, c_1 = 0.6$, and $c_2 = 0.1$. In particular, since $c_1 + c_2 < b$, mutual cooperation is by far the most abundant outcome.

As in the prisoner's dilemma, we can also use the stationary distribution to discuss the impact of selection strength on cooperation, see figure 3. If selection is weak, we can approximate the stationary distribution by (18), yielding

$$\mathbf{u} \approx \left(\frac{1}{4} + \frac{3b-2c_1-2c_2}{16} \cdot \beta, \frac{1}{4} + \frac{-b-2c_1+2c_2}{16} \cdot \beta, \frac{1}{4} + \frac{-b+2c_1-2c_2}{16} \cdot \beta, \frac{1}{4} + \frac{-b+2c_1+2c_2}{8} \cdot \beta \right). \tag{26}$$

In particular, the second state (C, D) in which only the high-cost player cooperates is always disfavored by selection. The other three states may be favored or disfavored, depending on the exact values of the benefits



and costs. For the strong selection limit, we obtain

$$\lim_{\beta \rightarrow \infty} \mathbf{u} = \begin{cases} (1, 0, 0, 0) & \text{if } b > c_1 + c_2 \\ (1/2, 0, 0, 1/2) & \text{if } b = c_1 + c_2 \\ (0, 0, 0, 1) & \text{if } b < c_1 + c_2. \end{cases} \quad (27)$$

That is, in the limit of strong selection, introspection dynamics always selects the risk-dominant equilibrium (even though mutual cooperation is payoff-dominant for all parameter values).

4.1.3. Volunteer's dilemma

As a final example of a social dilemma, we consider the volunteer's dilemma [75]. Here, already one cooperating player is sufficient for both players to get a benefit. The payoff matrix is

| | | |
|---|--------------------|--------------|
| | C | D |
| C | $b - c_1, b - c_2$ | $b - c_1, b$ |
| D | $b, b - c_2$ | $0, 0$ |

(28)

where $b > c_1 \geq c_2 > 0$. The game has two pure equilibria in which either only the first player or only the second player cooperates, (C, D) and (D, C). In both cases, the dilemma arises because players prefer the other player to volunteer as a cooperator.

For the volunteer's dilemma, we set $A = -c_1$, $A' = -c_2$, $B = -(b - c_1)$, $B' = -(b - c_2)$. By (17), the stationary distribution is

$$\mathbf{u} = \frac{1}{e^{\beta b} + e^{\beta(b+c_1)} + e^{\beta(b+c_2)} + e^{\beta(c_1+c_2)}} \cdot (e^{\beta b}, e^{\beta(b+c_2)}, e^{\beta(b+c_1)}, e^{\beta(c_1+c_2)}). \quad (29)$$

Table 1. Summary of our results on asymmetric social dilemmas with two strategies.

| Game | Prisoner's dilemma | Stag-hunt game | Volunteer's dilemma |
|---|---|--|--|
| Parameters (A, A', B, B') | $(-c_1, -c_2, c_1, c_2)$ | $(b - c_1, b - c_2, c_1, c_2)$ | $(-c_1, -c_2, -(b - c_1), -(b - c_2))$ |
| Stationary distribution (up to a constant) | $(1, e^{c_2\beta}, e^{c_1\beta}, e^{(c_1+c_2)\beta})$ | $(e^{b\beta}, e^{c_2\beta}, e^{c_1\beta}, e^{(c_1+c_2)\beta})$ | $(e^{b\beta}, e^{(b+c_2)\beta}, e^{(b+c_1)\beta}, e^{(c_1+c_2)\beta})$ |
| Weak selection $(\beta \ll 1) \frac{1}{4}(1, 1, 1, 1) + \frac{\beta}{16}(\star)$ | $2(-c_1 - c_2, -c_1 + c_2, c_1 - c_2, c_1 + c_2)$ | $(3b - 2(c_1 + c_2), -b - 2(c_1 - c_2), -b + 2(c_1 - c_2), -b + 2(c_1 + c_2))$ | $(b - 2(c_1 + c_2), b - 2(c_1 - c_2), b + 2(c_1 - c_2), -3b + 2(c_1 + c_2))$ |
| Strong selection ($\beta \rightarrow \infty$) | $(0, 0, 0, 1)$ | $(\frac{1}{2}, 0, 0, \frac{1}{2})$ if $b = c_1 + c_2$ $(0, 0, 0, 1)$ if $b < c_1 + c_2$ | $(0, \frac{1}{2}, \frac{1}{2}, 0)$ if $c_1 = c_2$ $(0, 1, 0, 0)$ if $c_1 < c_2$ |
| Pure Nash equilibria | (D, D) | (C, C) and (D, D) | (C, D) and (D, C) |

Similar to the stag-hunt game, there are two possible orderings of the four states, depending on how the benefit b relates to the total costs $c_1 + c_2$ of cooperation,

$$\begin{aligned}
 u_{DD} \leq u_{CC} \leq u_{CD} \leq u_{DC} & \quad \text{if } b \geq c_1 + c_2 \\
 u_{CC} \leq u_{DD} \leq u_{CD} \leq u_{DC} & \quad \text{if } b \leq c_1 + c_2.
 \end{aligned}
 \tag{30}$$

In either case, the most likely outcome is that the low-cost player 2 cooperates, whereas the high-cost player 1 defects, as illustrated in figure 3.

The impact of selection strength can be discussed analogously to the previous cases (figure 3). When selection is weak, the stationary distribution simplifies to

$$\mathbf{u} \approx \left(\frac{1}{4} + \frac{b - 2c_1 - 2c_2}{16} \cdot \beta, \frac{1}{4} + \frac{b - 2c_1 + 2c_2}{16} \cdot \beta, \frac{1}{4} + \frac{b + 2c_1 - 2c_2}{16} \cdot \beta, \frac{1}{4} + \frac{-3b + 2c_1 + 2c_2}{8} \cdot \beta \right).
 \tag{31}$$

Here, the third state (D, C) is always favored by selection. The other three states may be favored or disfavored, depending on the magnitudes of b, c_1, c_2 . For strong selection and $c_1 > c_2$,

$$\lim_{\beta \rightarrow \infty} \mathbf{u} = (0, 0, 1, 0).
 \tag{32}$$

That is, the low-cost player volunteers with certainty.

We briefly summarize the key results of this section in table 1. For each of the three social dilemmas we considered, we describe (i) the respective game parameters, (ii) the resulting stationary distribution, (iii) the respective weak selection limit, (iv) the strong selection limit, and (v) the game's Nash equilibria.

4.2. Comparing introspection and imitation in symmetric games

While the previous section focused on asymmetric social dilemmas, introspection dynamics is equally applicable to symmetric games. In the special case that the game has two strategies only, the assumption of symmetry implies $A = A'$ and $B = B'$, and the payoff matrix (16) simplifies to

$$\begin{array}{c|cc}
 & C & D \\
 \hline
 C & A & 0 \\
 D & 0 & B
 \end{array}
 \tag{33}$$

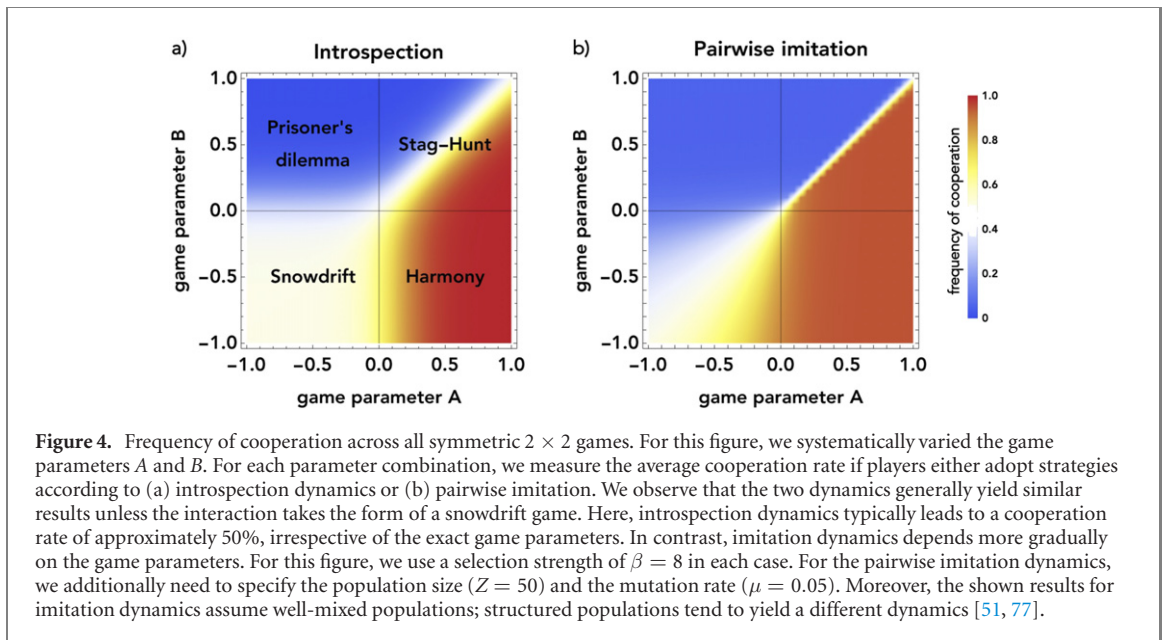
Similarly, the stationary distribution $\mathbf{u} = (u_{CC}, u_{CD}, u_{DC}, u_{DD})$ reduces to

$$\mathbf{u} = \frac{1}{2 + e^{A\beta} + e^{B\beta}} \cdot (e^{A\beta}, 1, 1, e^{B\beta}).
 \tag{34}$$

In particular, we can immediately see that $u_{CD} = u_{DC}$ for any A and B , as one would expect from a symmetric game. Moreover, the average cooperation probability of player 1 (and therefore also of player 2) becomes

$$\xi_C = u_{CC} + u_{CD} = \frac{1 + e^{A\beta}}{2 + e^{A\beta} + e^{B\beta}}.
 \tag{35}$$

For a given selection strength, this formula has only two free parameters. Hence, we can use this formula to explore the expected cooperation rate across all symmetric 2×2 games, by simultaneously varying both A and B as in figure 4(a). Depending on the signs of A and B , we recover the four classical symmetric games [76]: the prisoner's dilemma ($A < 0$ and $B > 0$), the stag-hunt game ($A > 0$ and $B > 0$), the snowdrift



game ($A < 0$ and $B < 0$), and the harmony game ($A > 0$ and $B < 0$). For each possible combination of (A, B), we plot the resulting cooperation probability according to (35). We observe that there are three qualitative regions: (i) for $B > 0$ and $A < B$, defection is either a dominant strategy, or it is risk-dominant. In this parameter region, we, therefore, observe comparably little cooperation. (ii) Conversely, for $A > 0$ and $B < A$, cooperation is either dominant or risk-dominant. As a consequence, introspection dynamics leads to almost full cooperation. (iii) If both $A < 0$ and $B < 0$ (i.e., in the snowdrift game), players have an incentive to choose the opposite strategy of the opponent. In this parameter region, we, therefore, observe an approximately equal share of cooperators and defectors.

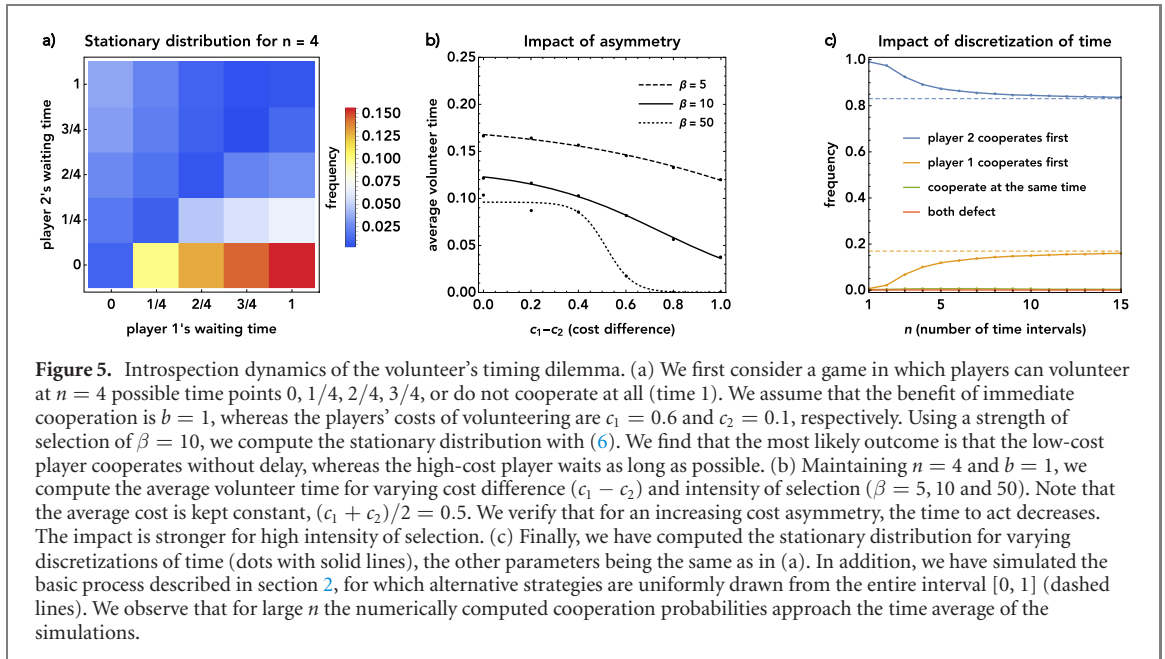
For symmetric games, we can compare introspection dynamics to the classical pairwise imitation rule [24], see figure 4(b) (for details of the implemented imitation model, see appendix D). In most parameter regions, the corresponding results are strikingly similar. Only in the snowdrift game regime, imitation leads to a more gradual change from full cooperation (when $B < 0$ and $A \approx 0$) to full defection (when $A < 0$ and $B \approx 0$). To explain this difference, we note that the imitation process takes place in an entire population of players. As a result, individuals do not adapt their strategy to a specific opponent, but rather to the population average. The resulting imitation dynamics can lead to a mixed equilibrium in which cooperators and defectors coexist. The exact position of this equilibrium changes gradually in the game parameters, as observed in figure 4(b). Overall, these results suggest that the two dynamics are comparable when the game tends to converge toward a homogeneous population. However, if the game has at least one stable mixed equilibrium (as in the snowdrift game), the predictions of the two models may diverge.

4.3. The volunteer's timing dilemma

In the previous examples, we have considered social dilemmas with two strategies only. In this section, we illustrate how introspection dynamics can be applied to a game with arbitrarily many strategies. To this end, we take the volunteer's dilemma and turn it into a timing dilemma. Here, players no longer only determine whether or not to volunteer. Instead, the game takes place over time, and players determine how long they wait for the co-player to volunteer before they volunteer themselves. This kind of game has been first studied by Weesie [78] as a model to explore the emergence of 'wait and see' behaviors in the context of voluntary action. Here we explore the game's introspection dynamics.

To formalize the game, we assume that within a given time interval $[t_0, t_{\max}] := [0, 1]$, players need to make a decision whether to volunteer. For simplicity, we assume that time is discretized, such that there are $n + 1$ evenly spaced time points $\{t_0, t_1, \dots, t_n\} = \{0, 1/n, \dots, 1\}$, at which players may volunteer. As before, we assume that players may have different costs to volunteer and that player 1's cost tends to be larger, $c_1 \geq c_2$. If at least one player volunteers during the time interval, both players derive some benefit. However, to add a component of time pressure, we assume that the benefit of cooperation is linearly degrading in time. That is, if one of the players cooperates immediately at time t_0 , both players get a benefit of $b > c_i$. However, if the first player to volunteer cooperates at time t_{\max} , the resulting benefit is zero.

A strategy for the volunteer's timing dilemma is now a rule that tells the player at which point to volunteer. To this end, we associate each strategy $\{S_0, S_1, \dots, S_n\}$ with a waiting time $\{t_0, t_1, \dots, t_n\}$. For



$i < n$, a player with strategy S_i volunteers at time t_i , unless the co-player has already volunteered earlier (in which case the focal player's cooperation is no longer required). For $i = n$, we associate the respective strategy S_n with not cooperating at all. If player 1 and player 2 adopt the strategies S_i and S'_j , respectively, the resulting payoffs are

$$\pi_{ij} = \begin{cases} (1 - \min[t_i, t_j]) \cdot b - c_1 & \text{if } i \leq j \text{ and } i < n \\ (1 - \min[t_i, t_j]) \cdot b & \text{otherwise} \end{cases} \quad (36)$$

$$\pi'_{ij} = \begin{cases} (1 - \min[t_i, t_j]) \cdot b - c_2 & \text{if } j \leq i \text{ and } j < n \\ (1 - \min[t_i, t_j]) \cdot b & \text{otherwise.} \end{cases}$$

Equivalently, the game can be represented by the payoff matrix

| | S'_0 | S'_1 | \dots | S'_n |
|----------|--------------------|--|----------|--|
| S_0 | $b - c_1, b - c_2$ | $b - c_1, b$ | \dots | $b - c_1, b$ |
| S_1 | $b, b - c_2$ | $\frac{n-1}{n}b - c_1, \frac{n-1}{n}b - c_2$ | \dots | $\frac{n-1}{n}b - c_1, \frac{n-1}{n}b$ |
| \vdots | \vdots | \vdots | \ddots | \vdots |
| S_n | $b, b - c_2$ | $\frac{n-1}{n}b, \frac{n-1}{n}b - c_2$ | \dots | $0, 0$ |

(37)

Note that for $n = 1$, we recover the payoffs of the original volunteer's dilemma.

We explore the introspection dynamics of the volunteer's timing dilemma numerically, by computing the stationary distribution with (6). To this end, we first consider a case in which $n = 4$, such that the players' possible waiting times are $t \in \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$. Moreover, we consider a normalized benefit of $b = 1$, and the player's cooperation costs are $c_1 = 0.7$ and $c_2 = 0.3$, respectively. The resulting stationary distribution is displayed in figure 5(a). As one may expect, we observe that the low-cost player is more cooperative; however, more surprisingly, this player typically cooperates without any delay.

A natural question, then, is if the cost asymmetry helps to solve the timing dilemma. In figure 5(b), we show that the higher the difference in costs, the faster one of the players volunteers. This positive effect of asymmetry is particularly pronounced when selection is strong. To explore how these results depend on our discretization of the time interval, we have repeated the above analysis for different $n \in \{1, \dots, 15\}$. In addition, we have run simulations in which players are able to volunteer at any time in $[0, 1]$. The results of this analysis are displayed in figure 5(c). Interestingly, the low-cost player is most likely to volunteer when $n = 1$ (the original volunteer's dilemma), in which case a 'wait and see' approach is ruled out by the design

of the game. However, across all values of n considered, the low-cost player always cooperates with a probability of at least 80%, implying that this player remains the most reliable volunteer.

5. Discussion and conclusion

Herein, we present a simple model of learning in social interactions. The model considers individuals who can choose among several strategies using introspection, that is, by reasoning about their strategies' prospective consequences. Compared to imitation models [24], this approach has the advantage that it can be applied to symmetric and asymmetric games alike. As another advantage, we can derive explicit formulas that describe how the system evolves in time, and what the long-run abundance of each strategy is. These formulas become particularly simple when players can only choose among two strategies, or when selection is weak [79–81].

Mathematically, the model takes the form of a Markov chain, whose states are the possible combinations of the strategies of the two players. In particular, if the players can choose among m and n strategies, respectively, there are mn states, arguably the minimum number of states that any learning model for such asymmetric games must have. While a similar Markov approach can also be used to analyze imitation processes in populations of players, the two approaches differ in their computational complexity. Population models need to record how many players apply any given strategy at any point in time. As a result, if a symmetric game with n strategies is played in a population of size Z , there are $\binom{Z+n-1}{n-1}$ possible states [82]. Since this number of states increases exponentially in n , numerically exact results are only feasible in games with a few strategies [83], or when mutations are rare [84–87]. In contrast, the computational complexity of introspection dynamics only depends polynomially on the number of strategies. As a result, the stationary distribution can be easily computed even in complex games with many possible moves [67].

Despite these differences in terms of computational complexity, the results of introspection dynamics are often in remarkable agreement with other evolutionary processes. For example, in the limit of weak selection, the introspection dynamics of two adapting learners becomes equivalent to a birth–death model of two co-evolving populations [60] (section 3.2). Similarly, across all symmetric 2×2 games, we find that introspection dynamics often recovers the results of pairwise imitation [24] (figure 4). The only notable exception occurs for the snowdrift game (as in [10]). In the snowdrift game, pairwise imitation typically selects for the mixed symmetric equilibrium, in which cooperators and defectors coexist. In contrast, introspection dynamics selects pure but asymmetric equilibria in which one player cooperates and the other defects.

To illustrate our analytical results, we apply our framework to a number of asymmetric games. As one particular example, we study the dynamics of the volunteer's timing dilemma [78]. In this game, one player is required to volunteer as quickly as possible to create a benefit for the whole group; yet each player may be tempted to wait, hoping the other player would give in first. When players differ in their costs to volunteer, we observe not only that the player with the lower cost is more likely to volunteer; we also note that this player usually volunteers without any delay (figure 5). This game thus illustrates how players may sometimes benefit from asymmetry because it helps them to coordinate more efficiently. Herein, we have studied this advantage of asymmetry by assuming that players have different costs. However, similar results could be obtained for other sources of asymmetry, such as when people differ in their endowments [43, 44], their productivities [54], or their strategic options more generally [41].

Here, we consider a comparably simple setup of introspective learning: there are two individuals who continually interact with each other. However, one could also imagine an entire population of introspective learners who interact with one another in such pairwise games. The results of such a process are likely to depend on the population's network topology [51, 77, 88–90]. Alternatively, one may also imagine that introspective learners engage in interactions that involve more than two players at a time [91, 92]. The introspection dynamics of such multiplayer games can easily be studied with simulations [54], similar to the 2-player games considered in this paper. Especially for asymmetric games with many unequal players, introspection dynamics can serve as a simple model to study the resulting learning processes.

Acknowledgments

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 the members of the Research Group Dynamics of Social Behavior and the Thesis Advisory Committee of MC for valuable feedback.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Appendix A. General properties of introspection dynamics

Here, we derive some properties of the transition matrix T defined in (3) in the main text and its stationary distribution \mathbf{u} .

Proposition 1 (Properties of the transition matrix T).

- (a) For a fixed j , T is unchanged by adding an arbitrary constant d_j to all payoffs π_{ij} .
- (b) For a fixed i , T is unchanged by adding an arbitrary constant d'_i to all payoffs π'_{ij} .
- (c) If β is finite, T is primitive.

Proof.

- (a) Transition probabilities in (3) in the main text only depend on payoff differences. If we fix j and replace all π_{ij} with $\tilde{\pi}_{ij} := \pi_{ij} + d_j$, the transition probabilities are unaltered since, for all $p \in \{1, \dots, m\}$, $\varphi_\beta(\tilde{\pi}_{pj} - \tilde{\pi}_{ij}) = \varphi_\beta((\pi_{pj} + d_j) - (\pi_{ij} + d_j)) = \varphi_\beta(\pi_{pj} - \pi_{ij})$.
- (b) A similar argument as in point (a) applies by fixing i and replacing all π'_{ij} with $\pi'_{ij} + d'_i$.
- (c) T is primitive if there is $a \in \mathbb{N}$ such that T^a is positive. Let $a = 2$. For arbitrary i, j, k, l ,

$$(T^2)_{ij,kl} = \sum_{p,q} T_{ij,pq} T_{pq,kl} \geq T_{ij,il} T_{il,kl} + T_{ij,kj} T_{kj,kl}.$$

For finite β , the entries of T as defined in (3) in the main text imply that both terms on the right-hand side of the above inequality are positive. Therefore, T^2 is positive and T is primitive under finite selection intensity. □

The first two properties described in proposition 1 are useful as they allow us to simplify the payoff matrices we need to consider (as illustrated in section 3.3). The last property helps us to analyze the long-term abundances of each strategy. Because T is primitive for finite β , the Perron–Frobenius theorem implies that the likelihood u_{ij} to observe the two players in a given state $(\mathbf{S}_i, \mathbf{S}'_j)$ converges as a function of time. Moreover, this likelihood can be determined by finding the unique solution of (5) in the main text

$$\begin{aligned} \mathbf{u} &= \mathbf{u} T \\ \mathbf{u} \mathbf{e}^\top &= 1. \end{aligned} \tag{A1}$$

While (A1) provides an implicit characterization of the stationary distribution \mathbf{u} , we can also provide an explicit representation.

Proposition 2 (An explicit representation of the stationary distribution). For $p \in \mathbb{N}$, let T be a row-stochastic and primitive $p \times p$ matrix, I denote the $p \times p$ identity matrix, and $U = \mathbf{e}^\top \mathbf{e}$ denote the $p \times p$ matrix whose entries are all equal to 1. Then $(I + U - T)$ is invertible, and the unique solution of (A1) can be given by $\mathbf{u} = \mathbf{e} (I + U - T)^{-1}$.

Proof. One proof of this result can be found in [70], where $(I + U - T)^{-1}$ is shown to be a fundamental matrix of the ergodic chain. In the following, we provide an independent proof.

Suppose \mathbf{u} satisfies (A1). If we multiply the second equation in (A1) by \mathbf{e} from the right, and subtract the result from the first equation, we obtain, after rearranging some of the terms

$$\mathbf{u} (I + U - T) = \mathbf{e}. \tag{A2}$$

Thus, for the proposition to hold, we only need to verify that $(I + U - T)$ is invertible. To this end, let $\lambda_1, \dots, \lambda_p$ be the eigenvalues of T and $\mathbf{w}_1, \dots, \mathbf{w}_p$ the corresponding right eigenvectors. Since T is row-stochastic it has an eigenvalue $\lambda_1 = 1$ with corresponding right eigenvector $\mathbf{w}_1 = (1, 1, \dots, 1)^\top$. By T 's primitivity, it follows that $I - T$ has a unique eigenvalue equal to 0 with corresponding right eigenvector \mathbf{w}_1 , while $1 - \lambda_k \neq 0$ for $k = 2, \dots, p$. Because $U = \mathbf{w}_1 \mathbf{w}_1^\top$, it follows by Brauer's theorem (example 1.2.8, p 51 of [93]), that the eigenvalues of $(U + I - T)$ are $\mathbf{w}_1^\top \mathbf{w}_1, 1 - \lambda_2, \dots, 1 - \lambda_p$. That is, in general the eigenvalues are the same as for the matrix $I - T$; only the eigenvalue corresponding to \mathbf{w}_1 gets replaced by

$\mathbf{w}_1^\top \mathbf{w}_1$. Since $\mathbf{w}_1^\top \mathbf{w}_1 = p > 0$ and $1 - \lambda_k \neq 0$ for $k = 2, \dots, p$, the matrix $(U + I - T)$ has no eigenvalue equal to 0. Therefore, it is invertible. \square

While the stationary distribution \mathbf{u} is well-defined for any finite β , in section 4.1 we also study which strategies are played as β approaches infinity. To this end, we first compute an expression for $\mathbf{u}(\beta)$ that is valid for finite β . Thereafter, we take the limit $\lim_{\beta \rightarrow \infty} \mathbf{u}(\beta)$ in \mathbb{R}^{mn} . For all considered 2×2 games, this yields a unique prediction of the strong selection limit—even for those games for which the respective limiting transition matrix $\lim_{\beta \rightarrow \infty} T(\beta)$ allows for several absorbing states. One example is the stag-hunt game (see table 1). Here, introspection dynamics predicts that if (\mathbf{C}, \mathbf{C}) is risk-dominant, players coordinate on mutual cooperation. Similarly, if (\mathbf{D}, \mathbf{D}) is risk-dominant, players coordinate on mutual defection. In comparison, according to the limiting transition matrix $\lim_{\beta \rightarrow \infty} T(\beta)$, both (\mathbf{C}, \mathbf{C}) and (\mathbf{D}, \mathbf{D}) are absorbing, irrespective of which one is risk-dominant. These results suggest that the strong selection limit of introspection dynamics can serve as an equilibrium selection device for 2×2 games, similar to other evolutionary dynamics [94, 95].

Appendix B. Introspection dynamics under weak selection

Here, we derive an explicit expression for the linear approximation of the stationary distribution when selection is weak, $\mathbf{u} \approx \mathbf{u}_0 + \mathbf{u}_1 \beta$. As we show in the main text, in (9) and (10), the constant term \mathbf{u}_0 of the stationary distribution obeys

$$\begin{aligned} \mathbf{u}_0 &= \mathbf{u}_0 T_0 \\ \mathbf{u}_0 \mathbf{e}^\top &= \mathbf{1}, \end{aligned} \tag{B1}$$

whereas the linear \mathbf{u}_1 term satisfies

$$\begin{aligned} \mathbf{u}_1 (I - T_0) &= \mathbf{u}_0 T_1 \\ \mathbf{u}_1 \mathbf{e}^\top &= 0. \end{aligned} \tag{B2}$$

Both systems can be solved explicitly. For that, we first compute $T_0 := T|_{\beta=0}$ and $T_1 := \partial T / \partial \beta|_{\beta=0}$, which are the constant and the linear term of the transition matrix. Based on (3), we obtain

$$(T_0)_{ij,kl} = \begin{cases} \frac{1}{4(m-1)} & \text{if } i \neq k \wedge j = l \\ \frac{1}{4(n-1)} & \text{if } i = k \wedge j \neq l \\ 0 & \text{if } i \neq k \wedge j \neq l \\ \frac{1}{2} & \text{if } i = k \wedge j = l \end{cases} \quad (T_1)_{ij,kl} = \begin{cases} \frac{\pi_{kj} - \pi_{ij}}{8(m-1)} & \text{if } i \neq k \wedge j = l \\ \frac{\pi'_{il} - \pi'_{ij}}{8(n-1)} & \text{if } i = k \wedge j \neq l \\ 0 & \text{if } i \neq k \wedge j \neq l \\ \sum_{p=1}^m \frac{\pi_{ij} - \pi_{pj}}{8(m-1)} + \sum_{q=1}^n \frac{\pi'_{ij} - \pi'_{iq}}{8(n-1)} & \text{if } i = k \wedge j = l. \end{cases} \tag{B3}$$

In the special case that both players have the same number of strategies, the stationary distribution takes a particularly simple form, as the following result shows.

Proposition 3 (Weak selection approximation of the invariant distribution).

(a) The system of equations (B1) has a unique solution given by

$$(\hat{\mathbf{u}}_0)_{ij} = \frac{1}{mn}. \tag{B4}$$

(b) The system of equations (B2) has a unique solution. For $m = n$, it takes the simple form

$$(\hat{\mathbf{u}}_1)_{ij} = \frac{1}{2n^4} \sum_{p=1}^n \sum_{q=1}^n ((\pi_{ij} + \pi_{iq} - \pi_{pj} - \pi_{pq}) + (\pi'_{ij} + \pi'_{pj} - \pi'_{iq} - \pi'_{pq})). \tag{B5}$$

Proof.

(a) To prove the first part, we note that T_0 is symmetric; indeed, swapping i with k and j with l in the expression for T_0 in (B3) shows that $(T_0)_{ij,kl} = (T_0)_{kl,ij}$. The stationary distribution of a Markov chain with symmetric transition matrix is uniform [70]. Since T is of size mn , the result in (B4) immediately follows.

- (b) To show that any solution to (B2) must be unique, we multiply the second equation in (B2) by the row-vector \mathbf{e} from the right, and add the result to the first equation. This yields

$$\mathbf{u}_1 (I + U - T_0) = \mathbf{u}_0 T_1, \tag{B6}$$

where again $U = \mathbf{e}^\top \mathbf{e}$. By proposition 2, the matrix $(I + U - T_0)$ is invertible, and hence any solution to (B2) is uniquely determined by $\mathbf{u}_1 = \mathbf{u}_0 T_1 (I + U - T_0)^{-1}$.

It remains to show that for $m = n$, the vector $\hat{\mathbf{u}}_1$ as defined by (B5) is in fact a solution of (B2). To this end, we first compute the right-hand side of the first condition in (B2) by plugging in the definitions of $\hat{\mathbf{u}}_0$ according to (B4) and of T_1 according to (B3),

$$\begin{aligned} (\hat{\mathbf{u}}_0 T_1)_{kl} &= \sum_{i=1}^n \sum_{j=1}^n (\hat{\mathbf{u}}_0)_{ij} (T_1)_{ij,kl} \stackrel{(B4)}{=} \sum_{i=1}^n \sum_{j=1}^n \frac{1}{n^2} (T_1)_{ij,kl} \\ &\stackrel{(B3)}{=} \frac{1}{n^2} \left(\sum_{i \neq k} \frac{\pi_{kl} - \pi_{il}}{8(n-1)} + \sum_{j \neq l} \frac{\pi'_{kl} - \pi'_{kj}}{8(n-1)} + \sum_{p=1}^n \frac{\pi_{kl} - \pi_{pl}}{8(n-1)} + \sum_{q=1}^n \frac{\pi'_{kl} - \pi'_{kq}}{8(n-1)} \right) \\ &= \frac{1}{n^2} \left(\sum_{p=1}^n \frac{\pi_{kl} - \pi_{pl}}{8(n-1)} + \sum_{p=1}^n \frac{\pi'_{kl} - \pi'_{kp}}{8(n-1)} + \sum_{p=1}^n \frac{\pi_{kl} - \pi_{pl}}{8(n-1)} + \sum_{p=1}^n \frac{\pi'_{kl} - \pi'_{kp}}{8(n-1)} \right) \\ &= \frac{1}{4(n-1)n^2} \sum_{p=1}^n ((\pi_{kl} - \pi_{pl}) + (\pi'_{kl} - \pi'_{kp})). \end{aligned} \tag{B7}$$

Similarly, we can compute the left-hand side of the first condition in (B2). By setting $m = n$ and using the definition of T_0 according to (B3), we obtain

$$\begin{aligned} (\hat{\mathbf{u}}_1 (I - T_0))_{kl} &= \sum_{i=1}^n \sum_{j=1}^n (\hat{\mathbf{u}}_1)_{ij} (I - T_0)_{ij,kl} \\ &\stackrel{(B3)}{=} \frac{(\hat{\mathbf{u}}_1)_{kl}}{2} - \sum_{i \neq k} \frac{(\hat{\mathbf{u}}_1)_{il}}{4(n-1)} - \sum_{j \neq l} \frac{(\hat{\mathbf{u}}_1)_{kj}}{4(n-1)} \\ &= \frac{(\hat{\mathbf{u}}_1)_{kl}}{2} - \left(\sum_{i=1}^n \frac{(\hat{\mathbf{u}}_1)_{il}}{4(n-1)} - \frac{(\hat{\mathbf{u}}_1)_{kl}}{4(n-1)} \right) - \left(\sum_{j=1}^n \frac{(\hat{\mathbf{u}}_1)_{kj}}{4(n-1)} - \frac{(\hat{\mathbf{u}}_1)_{kl}}{4(n-1)} \right) \\ &= \frac{n}{2(n-1)} (\hat{\mathbf{u}}_1)_{kl} - \sum_{i=1}^n \frac{(\hat{\mathbf{u}}_1)_{il}}{4(n-1)} - \sum_{j=1}^n \frac{(\hat{\mathbf{u}}_1)_{kj}}{4(n-1)}. \end{aligned} \tag{B8}$$

In a next step, we plug in the definition of $\hat{\mathbf{u}}_1$ according to (B5), yielding

$$\begin{aligned} (\hat{\mathbf{u}}_1 (I - T_0))_{kl} &= \frac{1}{4(n-1)n^3} \sum_{p=1}^n \sum_{q=1}^n ((\pi_{kl} + \pi_{kq} - \pi_{pl} - \pi_{pq}) + (\pi'_{kl} + \pi'_{pl} - \pi'_{kq} - \pi'_{pq})) \\ &\quad - \frac{1}{8(n-1)n^4} \sum_{i=1}^n \sum_{p=1}^n \sum_{q=1}^n ((\pi_{il} + \pi_{iq} - \pi_{pl} - \pi_{pq}) + (\pi'_{il} + \pi'_{pl} - \pi'_{iq} - \pi'_{pq})) \\ &\quad - \frac{1}{8(n-1)n^4} \sum_{j=1}^n \sum_{p=1}^n \sum_{q=1}^n ((\pi_{kj} + \pi_{kq} - \pi_{pj} - \pi_{pq}) + (\pi'_{kj} + \pi'_{pj} - \pi'_{kq} - \pi'_{pq})). \end{aligned} \tag{B9}$$

We can further simplify this expression by rewriting the sums in the second and third line,

$$\begin{aligned} (\hat{\mathbf{u}}_1 (I - T_0))_{kl} &= \frac{1}{4(n-1)n^3} \sum_{p=1}^n \sum_{q=1}^n ((\pi_{kl} + \pi_{kq} - \pi_{pl} - \pi_{pq}) + (\pi'_{kl} + \pi'_{pl} - \pi'_{kq} - \pi'_{pq})) \\ &\quad - \frac{1}{8(n-1)n^3} \sum_{p=1}^n \sum_{q=1}^n ((\pi_{pl} + \pi_{pq} - \pi_{pl} - \pi_{pq}) + (\pi'_{pl} + \pi'_{pl} - \pi'_{pq} - \pi'_{pq})) \\ &\quad - \frac{1}{8(n-1)n^3} \sum_{p=1}^n \sum_{q=1}^n ((\pi_{kq} + \pi_{kq} - \pi_{pq} - \pi_{pq}) + (\pi'_{kq} + \pi'_{pq} - \pi'_{kq} - \pi'_{pq})). \end{aligned} \tag{B10}$$

Now, the second and the third sum can be further simplified naturally. Then, by collecting like terms across the three sums, we eventually obtain a simple result,

$$\begin{aligned}
 (\hat{\mathbf{u}}_1(I-T_0))_{kl} &= \frac{1}{4(n-1)n^3} \sum_{p=1}^n \sum_{q=1}^n ((\pi_{kl} + \pi_{kq} - \pi_{pl} - \pi_{pq}) + (\pi'_{kl} + \pi'_{pl} - \pi'_{kq} - \pi'_{pq})) \\
 &\quad - \frac{1}{4(n-1)n^3} \sum_{p=1}^n \sum_{q=1}^n (\pi'_{pl} - \pi'_{pq}) - \frac{1}{4(n-1)n^3} \sum_{p=1}^n \sum_{q=1}^n (\pi_{kq} - \pi_{pq}) \\
 &= \frac{1}{4(n-1)n^3} \sum_{p=1}^n \sum_{q=1}^n ((\pi_{kl} - \pi_{pl}) + (\pi'_{kl} - \pi'_{kq})) \\
 &= \frac{1}{4(n-1)n^2} \sum_{p=1}^n (\pi_{kl} - \pi_{pl}) + (\pi'_{kl} - \pi'_{kp}). \tag{B11}
 \end{aligned}$$

This expression for $(\hat{\mathbf{u}}_1(I-T_0))_{kl}$ coincides with the expression we obtained for $(\hat{\mathbf{u}}_0 T_1)_{kl}$ in (B7). Thus, the first condition in (B2) is satisfied. To verify the second condition in (B2), we compute

$$\begin{aligned}
 (\hat{\mathbf{u}}_1) \cdot \mathbf{e}^\top &= \frac{1}{2n^4} \sum_{i=1}^n \sum_{j=1}^n \sum_{p=1}^n \sum_{q=1}^n ((\pi_{ij} + \pi_{iq} - \pi_{pj} - \pi_{pq}) + (\pi'_{ij} + \pi'_{pj} - \pi'_{iq} - \pi'_{pq})) \\
 &= \frac{1}{2n^2} \sum_{p=1}^n \sum_{q=1}^n ((\pi_{pq} + \pi_{pq} - \pi_{pq} - \pi_{pq}) + (\pi'_{pq} + \pi'_{pq} - \pi'_{pq} - \pi'_{pq})) = 0. \tag{B12}
 \end{aligned}$$

□

Next, we derive (14) in the main text. For a given stationary distribution \mathbf{u} , we define the respective marginal distributions by $\xi_i := \sum_{j=1}^n (\mathbf{u})_{ij}$ and $\xi'_j := \sum_{i=1}^m (\mathbf{u})_{ij}$. Using proposition 3 we can derive the following weak selection formulas for ξ_i and ξ'_j when the two players have the same number of strategies ($m = n$).

Corollary 1 (Marginal distributions for weak selection). For $m = n$ and small β , the abundance of the strategies \mathbf{S}_i and \mathbf{S}'_j is

$$\begin{aligned}
 \xi_i &= \frac{1}{n} + \frac{\beta}{n^3} \sum_{p=1}^n \sum_{q=1}^n (\pi_{iq} - \pi_{pq}) + \mathcal{O}(\beta^2), \\
 \xi'_j &= \frac{1}{n} + \frac{\beta}{n^3} \sum_{p=1}^n \sum_{q=1}^n (\pi'_{pj} - \pi'_{pq}) + \mathcal{O}(\beta^2). \tag{B13}
 \end{aligned}$$

Proof. To obtain the first equation, we use the weak selection formula in proposition 3 and sum up over all co-player's strategies \mathbf{S}'_j ,

$$\begin{aligned}
 \xi_i &= \sum_{j=1}^n (\mathbf{u}_0)_{ij} + \left(\sum_{j=1}^n (\mathbf{u}_1)_{ij} \right) \beta + \mathcal{O}(\beta^2) \\
 &= \sum_{j=1}^n \frac{1}{n^2} + \frac{1}{2n^4} \sum_{j=1}^n \sum_{p=1}^n \sum_{q=1}^n ((\pi_{ij} + \pi_{iq} - \pi_{pj} - \pi_{pq}) + (\pi'_{ij} + \pi'_{pj} - \pi'_{iq} - \pi'_{pq})) \beta + \mathcal{O}(\beta^2) \\
 &= \frac{1}{n} + \frac{1}{2n^3} \sum_{p=1}^n \sum_{q=1}^n ((\pi_{iq} + \pi_{iq} - \pi_{pq} - \pi_{pq}) + (\pi'_{iq} + \pi'_{pq} - \pi'_{iq} - \pi'_{pq})) \beta + \mathcal{O}(\beta^2) \\
 &= \frac{1}{n} + \frac{\beta}{n^3} \sum_{p=1}^n \sum_{q=1}^n (\pi_{iq} - \pi_{pq}) + \mathcal{O}(\beta^2). \tag{B14}
 \end{aligned}$$

The analogous formula for ξ'_j follows by symmetry. □

Three remarks are in order.

- (a) **Independence of the co-player's payoffs.** According to corollary 1, the abundance of a player's strategy in the limit of weak selection only depends on that player's payoffs. For instance, in (B13), the

abundance ξ_i depends on the first player's payoffs π_{pq} , but it is independent of the second player's payoffs π'_{pq} . While such a result may appear intuitive, it is important to note that this result only holds in the limit of weak selection. As an example, consider the game with 2×2 payoff matrix

$$\begin{array}{c|cc}
 & \mathbf{S}'_1 & \mathbf{S}'_2 \\
 \hline
 \mathbf{S}_1 & 2, x & 0, 0 \\
 \mathbf{S}_2 & 0, x & 1, 0
 \end{array} \tag{B15}$$

For this game, (B13) implies that for sufficiently small β , the marginal abundance of \mathbf{S}_1 is approximately given by $\xi_1 = \frac{1}{2} + \beta \frac{3}{8}$, irrespective of the value of x . However, in the limit of strong selection, $\beta \rightarrow \infty$, one can show that $\xi_1 = 1$ if x is positive, but $\xi_1 = 0$ if x is negative. Therefore, the co-player's payoffs do in general affect how likely a player is to adopt a certain strategy. Only when selection is weak (such that each co-player's strategy is played with approximately equal frequency), this dependency disappears.

- (b) **Comparing the abundance of different strategies.** Corollary 1 also allows us to rank the different strategies of a player according to how often they are played in the stationary distribution. To this end, we say strategy \mathbf{S}_i is favored to \mathbf{S}_k , and we write $\mathbf{S}_i \succ \mathbf{S}_k$, if $\xi_i > \xi_k$. An analogous notation can be defined for the column player. By corollary 1 we find that when selection is weak, then $\mathbf{S}_i \succ \mathbf{S}_k$ if and only if $\sum_{q=1}^n \pi_{iq} > \sum_{q=1}^n \pi_{kq}$. Again, the respective condition only depends on the focal player's payoffs π_{pq} and not on the co-player's payoffs π'_{pq} . However, as before, this independence vanishes for stronger selection (which can be shown with the same example (B15)).
- (c) **Comparative statics.** Finally, we can use corollary 1 to compute how changes in the players' payoffs affect the resulting strategy abundances in the limit of weak selection. By (B13), we note that $\partial \xi_i / \partial \pi_{pq}$ is positive if $p = i$, while it is negative if $p \neq i$. Hence, increasing any one of the payoffs π_{iq} has a positive effect on the long-run abundance of strategy \mathbf{S}_i , whereas increasing any one of the other payoffs π_{pq} with $p \neq i$ has a negative effect.

Appendix C. Higher order approximations

In the previous section, we have derived a linear approximation for the long run abundance of each strategy profile $(\mathbf{S}_i, \mathbf{S}'_j)$ under weak selection. Using similar techniques, we can obtain arbitrary higher-order approximations. To this end, we define the k th derivative of \mathbf{v} and T , respectively, evaluated at $\beta = 0$,

$$\mathbf{u}_k := \left. \frac{\partial^k \mathbf{u}}{\partial \beta^k} \right|_{\beta=0} \quad \text{and} \quad T_k := \left. \frac{\partial^k T}{\partial \beta^k} \right|_{\beta=0}. \tag{C1}$$

By computing the respective higher-order expansion of (7) and comparing coefficients with respect to the power β^k , we obtain for $k \geq 1$

$$\begin{aligned}
 \mathbf{u}_k &= \sum_{j=0}^k \binom{k}{j} \mathbf{u}_j T_{k-j} \\
 \mathbf{u}_k \mathbf{e}^\top &= 0.
 \end{aligned} \tag{C2}$$

By the first equation, it follows that we can compute \mathbf{u}_k recursively from all lower terms \mathbf{u}_j with $j < k$ as

$$\mathbf{u}_k (I - T_0) = \sum_{j=0}^{k-1} \binom{k}{j} \mathbf{u}_j T_{k-j}. \tag{C3}$$

By multiplying the second equation in (C2) by the row-vector \mathbf{e} from the right, and adding the result to (C3), we obtain

$$\mathbf{u}_k (I + U - T_0) = \sum_{j=0}^{k-1} \binom{k}{j} \mathbf{u}_j T_{k-j}. \tag{C4}$$

Since $(I + U - T_0)$ is invertible by proposition 2, we obtain the explicit solution

$$\mathbf{u}_k = \left(\sum_{j=0}^{k-1} \binom{k}{j} \mathbf{u}_j T_{k-j} \right) \cdot (U + I - T_0)^{-1}. \tag{C5}$$

Table 2. The Eulerian numbers $E(k, l)$ for $1 \leq k \leq 5$ and $0 \leq l \leq 5$.

| $k \setminus l$ | 0 | 1 | 2 | 3 | 4 | 5 |
|-----------------|---|----|----|----|---|---|
| 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 2 | 1 | 1 | 0 | 0 | 0 | 0 |
| 3 | 1 | 4 | 1 | 0 | 0 | 0 |
| 4 | 1 | 11 | 11 | 1 | 0 | 0 |
| 5 | 1 | 26 | 66 | 26 | 1 | 0 |

For introspection dynamics, the first factor on the right-hand side can be further simplified. We note that T_i is the zero matrix for all even $i \geq 2$, as the following result shows.

Proposition 4. For T_i defined as in (C1), $T_i = 0$ for any even integer $i \geq 2$.

Proof. By its definition (3) in the main text, we can write all entries of the transition matrix T as a linear combination of constant terms and functions of the form $\varphi_\beta(\Delta\pi) = (1 + e^{-\beta\Delta\pi})^{-1} \cdot e^x$. Since, by taking the k th derivative of (2) in the main text,

$$\frac{\partial^k \varphi_\beta}{\partial \beta^k} = \frac{(-\Delta\pi)^k \cdot e^{-\beta\Delta\pi}}{(1 + e^{-\beta\Delta\pi})^{k+1}} \sum_{l=0}^k (-1)^{l+1} \cdot E(k, l) \cdot e^{-l\beta \cdot \Delta\pi}, \quad k \geq 1 \tag{C6}$$

the entries of T_k are proportional to

$$\left. \frac{\partial^k \varphi_\beta}{\partial \beta^k} \right|_{\beta=0} = \frac{-(-\Delta\pi)^k}{2^{k+1}} \sum_{l=0}^k (-1)^{l+1} \cdot E(k, l), \tag{C7}$$

where $E(k, l)$ are the Eulerian numbers defined by

$$E(k, l) = \sum_{r=0}^l (-1)^r \binom{k+1}{r} (l+1-r)^k.$$

For $1 \leq k \leq 5$ and $0 \leq l \leq 5$, these numbers are tabulated in table 2. A useful property of Eulerian numbers is the symmetry $E(k, l) = E(k, k-l-1)$. The sum in (C7) translates into adding the numbers of the k th line of the table with alternating signs. For even k , that sum is zero because the different Eulerian numbers exactly cancel each other out due to their symmetry. □

As a result, the first seven terms of the Taylor expansion of the stationary distribution \mathbf{u} are

$$\begin{aligned} \mathbf{u}_0 &= \mathbf{e} (I + U - T_0)^{-1} \\ \mathbf{u}_1 &= (\mathbf{u}_0 T_1) (I + U - T_0)^{-1} \\ \mathbf{u}_2 &= (2\mathbf{u}_1 T_1) (I + U - T_0)^{-1} \\ \mathbf{u}_3 &= (3\mathbf{u}_2 T_1 + \mathbf{u}_0 T_3) (I + U - T_0)^{-1} \\ \mathbf{u}_4 &= (4\mathbf{u}_3 T_1 + 4\mathbf{u}_1 T_3) (I + U - T_0)^{-1} \\ \mathbf{u}_5 &= (5\mathbf{u}_4 T_1 + 10\mathbf{u}_2 T_3 + \mathbf{u}_0 T_5) (I + U - T_0)^{-1} \\ \mathbf{u}_6 &= (6\mathbf{u}_5 T_1 + 20\mathbf{u}_3 T_3 + 6\mathbf{u}_1 T_5) (I + U - T_0)^{-1}. \end{aligned} \tag{C8}$$

Appendix D. Pairwise imitation dynamics

Pairwise imitation dynamics is a frequency-dependent update rule where players can imitate other players' strategies [24]. The dynamics takes place in an entire population of size Z . For symmetric games with two strategies, as considered in section 4.2, the state of the population can be defined by the number of individuals i who use strategy C (the number of individuals who use strategy D is then $Z - i$). Given the current population state i , one can calculate the expected payoffs of cooperators and defectors, respectively, as

$$\begin{aligned} f_C &= \frac{i-1}{Z-1} \pi_{CC} + \frac{Z-i}{Z-1} \pi_{CD} \\ f_D &= \frac{i}{Z-1} \pi_{DC} + \frac{Z-i-1}{Z-1} \pi_{DD}. \end{aligned} \tag{D1}$$

Similar to introspection dynamics, pairwise imitation dynamics assumes that at regular time intervals, a randomly chosen player is given the opportunity to revise their strategy. With probability μ (the mutation rate), this player simply adopts a randomly chosen strategy. With probability $1 - \mu$, the player instead picks a randomly chosen role model from the population. If the focal player's payoff is π and the role model's payoff is $\tilde{\pi}$, the focal player adopts the role model's strategy with probability $\varphi_\beta(\tilde{\pi} - \pi)$, where φ_β is again the Fermi function defined by (2).

This process can also be described by a Markov chain. Given that the current population state is i , the probability that a **D**-player changes to **C** in the next time step is

$$T_{i,i+1} = (1 - \mu) \left[\frac{Z-i}{Z} \frac{i}{Z-1} \frac{1}{1 + e^{-\beta(f_C - f_D)}} \right] + \mu \frac{Z-i}{Z}. \quad (\text{D2a})$$

Similarly, the probability that a **C**-player changes to **D** is

$$T_{i,i-1} = (1 - \mu) \left[\frac{i}{Z} \frac{Z-i}{Z-1} \frac{1}{1 + e^{-\beta(f_D - f_C)}} \right] + \mu \frac{i}{Z}. \quad (\text{D2b})$$

Since only one player is allowed to update at each time step,

$$T_{i,j} = 0 \quad \text{for all } j \notin \{i-1, i, i+1\}. \quad (\text{D2c})$$

Finally, for the normalization condition to hold, the probability of remaining in the same state is

$$T_{i,i} = 1 - \sum_{i' \neq i} T_{i,i'}. \quad (\text{D2d})$$

As a result, the transition matrix is

$$\begin{bmatrix} T_{0,0} & T_{0,1} & 0 & 0 & 0 & \dots & 0 & 0 \\ T_{1,0} & T_{1,1} & T_{1,2} & 0 & 0 & \dots & 0 & 0 \\ 0 & T_{2,1} & T_{2,2} & T_{2,3} & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & T_{Z,Z-1} & T_{Z,Z} \end{bmatrix}.$$

As before, we can characterize the long-run dynamics of this process by computing the stationary distribution $\mathbf{u} = (u_0, u_1, \dots, u_Z)$. In this case, however, u_i denotes the prevalence of a population state with i cooperators and $Z - i$ defectors.

To construct a metric that is comparable to the one used for introspection dynamics, we compute the average fraction of cooperation,

$$p_C = \sum_{i=0}^Z u_i \frac{i}{Z} \quad (\text{D3})$$

which we plot in figure 4(b). Additionally, one can also obtain the average probability of drawing **CC**, **CD**, **DC**, and **DD** pairs from the population

$$p_{CC} = \sum_{i=0}^Z u_i \frac{i}{Z} \frac{i-1}{Z-1} \quad (\text{D4a})$$

$$p_{CD} = p_{DC} = \sum_{i=0}^Z u_i \frac{i}{Z} \frac{Z-i}{Z-1} \quad (\text{D4b})$$

$$p_{DD} = \sum_{i=0}^Z u_i \frac{Z-i}{Z} \frac{Z-i-1}{Z-1}. \quad (\text{D4c})$$

Again, $p_{CD} = p_{DC}$ because players are indistinguishable in the symmetric case. These quantities are comparable to the state frequencies of our stationary distribution, $\mathbf{u} = (u_{CC}, u_{CD}, u_{DC}, u_{DD})$.

ORCID iDs

M C Couto  <https://orcid.org/0000-0002-2868-9685>

S Giaimo  <https://orcid.org/0000-0003-0421-3065>

C Hilbe  <https://orcid.org/0000-0001-5116-955X>

References

- [1] Nowak M A 2006 *Evolutionary Dynamics* (Cambridge, MA: Harvard University Press)
- [2] Sigmund K 2010 *The Calculus of Selfishness* (Princeton, NJ: Princeton University Press)
- [3] Huang W, Hauert C and Traulsen A 2015 Stochastic game dynamics under demographic fluctuations *Proc. Natl Acad. Sci. USA* **112** 9064–9
- [4] Gokhale C S and Hauert C 2016 Eco-evolutionary dynamics of social dilemmas *Theor. Popul. Biol.* **111** 28–42
- [5] Hilbe C, Šimsa Š, Chatterjee K and Nowak M A 2018 Evolution of cooperation in stochastic games *Nature* **559** 246–9
- [6] Sandholm T W and Crites R H 1996 Multiagent reinforcement learning in the iterated prisoner's dilemma *Biosystems* **37** 147–66
- [7] Fudenberg D and Levine D 1998 *The Theory of Learning in Games* (Cambridge, MA: MIT Press)
- [8] Macy M W and Flache A 2002 Learning dynamics in social dilemmas *Proc. Natl Acad. Sci. USA* **99** 7229–36
- [9] Tuyls K, Heytens D, Nowe A and Manderick B 2003 Extended replicator dynamics as a key to reinforcement learning in multi-agent systems *14th European Conference on Machine Learning* (Cavtat-Dubrovnik, Croatia) 2837 ed N Lavrač, D Gamberger, H Blockeel and L Blockeel (Berlin, Heidelberg: Springer) pp 421–31
- [10] Amaral M A and Javarone M A 2018 Heterogeneous update mechanisms in evolutionary games: mixing innovative and imitative dynamics *Phys. Rev. E* **97** 042305
- [11] Pereira L M and Santos F C 2019 Counterfactual thinking in cooperation dynamics *International conference on Model-Based Reasoning* (2018) 49 ed Á Nepomuceno-Fernández, L Magnani, F Salguero-Lamillar, C Barés-Gómez and M Fontaine (Berlin: Springer) pp 69–82
- [12] Santos F P, Levin S A and Vasconcelos V V 2021 Biased perceptions explain collective action deadlocks and suggest new mechanisms to prompt cooperation *iScience* **24** 102375
- [13] Pangallo M, Sanders J B, Galla T and Farmer J D 2022 Towards a taxonomy of learning dynamics in 2×2 games *Games Econ. Behav.* **132** 1–21
- [14] Maynard Smith J 1982 *Evolution and the Theory of Games* (Cambridge: Cambridge University Press)
- [15] Hofbauer J and Sigmund K 1998 *Evolutionary Games and Population Dynamics* (Cambridge: Cambridge University Press)
- [16] Traulsen A and Hauert C 2009 Stochastic evolutionary game dynamics *Reviews of Nonlinear Dynamics and Complexity* ed H G Schuster (Weinheim: Wiley) pp 25–61
- [17] Kerr B, Godfrey-Smith P and Feldman M W 2004 What is altruism? *Trends Ecol. Evol.* **19** 135–40
- [18] Nowak M A 2012 Evolving cooperation *J. Theor. Biol.* **299** 1–8
- [19] Fudenberg D and Tirole J 1998 *Game Theory* 6th edn (Cambridge, MA: MIT Press)
- [20] Roca C P, Cuesta J A and Sánchez A 2009 Promotion of cooperation on networks? The myopic best response case *Eur. Phys. J. B* **71** 587–95
- [21] Wu B, Bauer B, Galla T and Traulsen A 2015 Fitness-based models and pairwise comparison models of evolutionary games are typically different—even in unstructured populations *New J. Phys.* **17** 023043
- [22] Schlag K H 1998 Why imitate, and if so, how? *J. Econ. Theor.* **78** 130–56
- [23] Szabó G and Tóke C 1998 Evolutionary prisoner's dilemma game on a square lattice *Phys. Rev. E* **58** 69–73
- [24] Traulsen A, Nowak M A and Pacheco J M 2006 Stochastic dynamics of invasion and fixation *Phys. Rev. E* **74** 011909
- [25] Peña J, Volken H, Pestelacci E and Tomassini M 2009 Conformity hinders the evolution of cooperation on scale-free networks *Phys. Rev. E* **80** 016110
- [26] Szolnoki A, Wang Z and Perc M 2012 Wisdom of groups promotes cooperation in evolutionary social dilemmas *Sci. Rep.* **2** 576
- [27] Grujic J et al 2014 A comparative analysis of spatial prisoner's dilemma experiments: conditional cooperation and payoff irrelevance *Sci. Rep.* **4** 4615
- [28] Selten R 1998 Aspiration adaption theory *J. Math. Phys.* **42** 191–214
- [29] Posch M 1999 Win-stay, lose-shift strategies for repeated games—memory length, aspiration levels and noise *J. Theor. Biol.* **198** 183–95
- [30] Du J, Wu B, Altrock P M and Wang L 2014 Aspiration dynamics of multi-player games in finite populations *J. R. Soc. Interface* **11** 1742–5662
- [31] Matsui A 1992 Best response dynamics and socially stable strategies *J. Econ. Theor.* **57** 343–62
- [32] Blume L E 1995 The statistical mechanics of best-response strategy revision *Games Econ. Behav.* **11** 111–45
- [33] Berger U 2007 Brown's original fictitious play *J. Econ. Theor.* **135** 572–8
- [34] Szolnoki A and Chen X 2018 Competition and partnership between conformity and payoff-based imitations in social dilemmas *New J. Phys.* **20** 093008
- [35] Danku Z, Wang Z and Szolnoki A 2018 Imitate or innovate: competition of strategy updating attitudes in spatial social dilemma games *Europhys. Lett.* **121** 18002
- [36] Szolnoki A and Danku Z 2018 Dynamic-sensitive cooperation in the presence of multiple strategy updating rules *Physica A* **511** 371–7
- [37] Hofbauer J and Hopkins E 2005 Learning in perturbed asymmetric games *Games Econ. Behav.* **52** 133–52
- [38] McAvoy A and Hauert C 2015 Asymmetric evolutionary games *PLoS Comput. Biol.* **11** e1004349
- [39] Nowak M A, Page K M and Sigmund K 2000 Fairness versus reason in the ultimatum game *Science* **289** 1773–5
- [40] Rand D G, Tarnita C E, Ohtsuki H and Nowak M A 2013 Evolution of fairness in the one-shot anonymous ultimatum game *Proc. Natl Acad. Sci. USA* **110** 2581–6
- [41] Chatterjee K, Zufferey D and Nowak M A 2012 Evolutionary game dynamics in populations with different learners *J. Theor. Biol.* **301** 161–73
- [42] Hilbe C, Hagel K and Milinski M 2016 Asymmetric power boosts extortion in an economic experiment *PLoS One* **11** e0163867
- [43] Vasconcelos V V, Santos F C, Pacheco J M and Levin S A 2014 Climate policies under wealth inequality *Proc. Natl Acad. Sci. USA* **111** 2212–6
- [44] Abou Chakra M, Bumann S, Schenk H, Oschlies A and Traulsen A 2018 Immediate action is the best strategy when facing uncertain climate change *Nat. Commun.* **9** 2566
- [45] Chica M, Hernández J M and Gidumal J B 2021 A collective risk dilemma for tourism restrictions under the COVID-19 context *Sci. Rep.* **11** 5043
- [46] von Gerwen N, Buskens V and van der Lippe T 2018 Individual training and employees' cooperative behavior: evidence from a contextualized laboratory experiment *Ration. Soc.* **30** 432–62

- [47] Szolnoki A, Perc M and Danku Z 2008 Towards effective payoffs in the prisoner's dilemma game on scale-free networks *Physica A* **387** 2075–82
- [48] McAvoy A, Allen B and Nowak M A 2020 Social goods dilemmas in heterogeneous societies *Nat. Human Behav.* **4** 819–31
- [49] Antonioni A, Pereda M, Cronin K A, Tomassini M and Sánchez A 2018 Collaborative hierarchy maintains cooperation in asymmetric games *Sci. Rep.* **8** 5375
- [50] Sysi-Aho M, Saramäki J, Kertész J and Kaski K 2005 Spatial snowdrift game with myopic agents *Eur. Phys. J. B* **44** 129–35
- [51] Perc M, Jordan J J, Rand D G, Wang Z, Boccaletti S and Szolnoki A 2017 Statistical physics of human cooperation *Phys. Rep.* **687** 1–51
- [52] Szabó G and Szolnoki A 2012 Selfishness, fraternity, and other-regarding preference in spatial evolutionary games *J. Theor. Biol.* **299** 81–7
- [53] Szolnoki A and Perc M 2014 Evolution of extortion in structured populations *Phys. Rev. E* **89** 022804
- [54] Hauser O P, Hilbe C, Chatterjee K and Nowak M A 2019 Social dilemmas among unequals *Nature* **572** 524–7
- [55] McAvoy A, Kates-Harbeck J, Chatterjee K and Hilbe C 2021 Evolutionary (in)stability of selfish learning in repeated games (arXiv:2105.06199)
- [56] Schuster P and Sigmund K 1981 Coyness, philandering and stable strategies *Anim. Behav.* **29** 186–92
- [57] Gaunersdorfer A, Hofbauer J and Sigmund K 1991 On the dynamics of asymmetric games *Theor. Popul. Biol.* **39** 345–57
- [58] Hofbauer J 1996 Evolutionary dynamics for bimatrix games: a Hamiltonian system? *J. Math. Biology* **34** 675–88
- [59] Cremer J, Reichenbach T and Frey E 2008 Anomalous finite-size effects in the battle of the sexes *Eur. Phys. J. B* **63** 373–80
- [60] Ohtsuki H 2010 Stochastic evolutionary dynamics of bimatrix games *J. Theor. Biol.* **264** 136–42
- [61] Sekiguchi T and Ohtsuki H 2017 Fixation probabilities of strategies for bimatrix games in finite populations *Dyn. Games Appl.* **7** 93–111
- [62] Veller C, Hayward L K, Hilbe C and Nowak M A 2017 The red queen and king in finite populations *Proc. Natl Acad. Sci. USA* **114** E5396–405
- [63] Tuyls K et al 2018 Symmetric decomposition of asymmetric games *Sci. Rep.* **8** 1015
- [64] Hoffman M, Hilbe C and Nowak M A 2018 The signal-burying game can explain why we obscure positive traits and good deeds *Nat. Human Behav.* **2** 397–404
- [65] Broere J, Buskens V, Weesie J and Stoof H 2017 Network effects on coordination in asymmetric games *Sci. Rep.* **7** 17016
- [66] Correia A D, Leestmaker L L, Stoof H T C and Broere J J 2022 Asymmetric games on networks: towards an Ising-model representation *Physica A* **593** 126972
- [67] Galla T and Farmer J D 2013 Complex dynamics in learning complicated games *Proc. Natl Acad. Sci. USA* **110** 1232–6
- [68] Blume L E 1993 The statistical mechanics of strategic interaction *Games Econ. Behav.* **5** 387–424
- [69] Meyer C D 2000 *Matrix Analysis and Applied Linear Algebra* (Philadelphia, PA: SIAM)
- [70] Kemeny J G and Snell J L 1983 *Finite Markov Chains* (New York: Springer)
- [71] Wu B, Altrock P M, Wang L and Traulsen A 2010 Universality of weak selection *Phys. Rev. E* **82** 046106
- [72] Adlam B and Nowak M A 2014 Universality of fixation probabilities in randomly structured populations *Sci. Rep.* **4** 6692
- [73] Pacheco J M, Santos F C, Souza M O and Skyrms B 2008 Evolutionary dynamics of collective action in N -person stag-hunt dilemmas *Proc. R. Soc. B* **276** 315–21
- [74] Harsanyi J C and Selten R 1988 *A General Theory of Equilibrium Selection in Games* (Cambridge, MA: MIT Press)
- [75] Diekmann A 1985 Volunteer's dilemma *J. Conflict Resolut.* **29** 605–10
- [76] Martínez-Vaquero L A, Cuesta J A and Sánchez A 2012 Generosity pays in the presence of direct reciprocity: a comprehensive study of 2×2 repeated games *PLoS One* **7** E35135
- [77] Roca C P, Cuesta J A and Sánchez A 2009 Evolutionary game theory: temporal and spatial effects beyond replicator dynamics *Phys. Life Rev.* **6** 208–49
- [78] Weesie J 1993 Asymmetry and timing in the volunteer's dilemma *J. Conflict Resolut.* **37** 569–90
- [79] Nowak M A, Sasaki A, Taylor C and Fudenberg D 2004 Emergence of cooperation and evolutionary stability in finite populations *Nature* **428** 646–50
- [80] Antal T and Scheuring I 2006 Fixation of strategies for an evolutionary game in finite populations *Bull. Math. Biol.* **68** 1923–44
- [81] Wild G and Traulsen A 2007 The different limits of weak selection and the evolutionary dynamics of finite populations *J. Theor. Biol.* **247** 382–90
- [82] Vasconcelos V V, Santos F P, Santos F C and Pacheco J M 2017 Stochastic dynamics through hierarchically embedded Markov chains *Phys. Rev. Lett.* **118** 058301
- [83] Couto M C, Pacheco J M and Santos F C 2020 Governance of risky public goods under graduated punishment *J. Theor. Biol.* **505** 110423
- [84] Fudenberg D and Imhof L A 2006 Imitation processes with small mutations *J. Econ. Theor.* **131** 251–62
- [85] Wu B, Gokhale C S, Wang L and Traulsen A 2012 How small are small mutation rates? *J. Math. Biol.* **64** 803–27
- [86] McAvoy A 2015 Comment on 'Imitation processes with small mutations' *J. Econ. Theory* **159** 66–9
- [87] Veller C and Hayward L K 2016 Finite-population evolution with rare mutations in asymmetric games *J. Econ. Theor.* **162** 93–113
- [88] Santos F C and Pacheco J M 2005 Scale-free networks provide a unifying framework for the emergence of cooperation *Phys. Rev. Lett.* **95** 098104
- [89] Santos F C, Pacheco J M and Lenaerts T 2006 Evolutionary dynamics of social dilemmas in structured heterogeneous populations *Proc. Natl Acad. Sci. USA* **103** 3490–4
- [90] Ohtsuki H and Nowak M A 2006 The replicator equation on graphs *J. Theor. Biol.* **243** 86–97
- [91] Gokhale C S and Traulsen A 2010 Evolutionary games in the multiverse *Proc. Natl Acad. Sci. USA* **107** 5500–4
- [92] Perc M, Gómez-Gardeñes J, Szolnoki A, Floría L M and Moreno Y 2013 Evolutionary dynamics of group interactions on structured populations: a review *J. R. Soc. Interface* **10** 20120997
- [93] Horn R A and Johnson C R 1985 *Matrix Analysis* (Cambridge: Cambridge University Press)
- [94] Zhang B and Hofbauer J 2015 Equilibrium selection via replicator dynamics in 2×2 coordination games *Int. J. Game Theory* **44** 433–48
- [95] Zhang B and Hofbauer J 2016 Quantal response methods for equilibrium selection in 2×2 coordination games *Games Econ. Behav.* **97** 19–31