WORKING PAPERS

N° 153

Cooperative dilemmas with binary actions and multiple players

Jorge Peña and Georg Nöldeke

Institute for Advanced Study in Toulouse

Cooperative dilemmas with binary actions and multiple players

Jorge Peña[*†] Georg Nöldeke[‡]

February 27, 2023

Abstract

The prisoners' dilemma, the snowdrift game, and the stag hunt are simple two-player games that are often considered as prototypical examples of cooperative dilemmas across disciplines. However, surprisingly little consensus exists about the precise mathematical meaning of the words "cooperation" and "cooperative dilemma" for these and other binary-action games, in particular when considering interactions among more than two players. Here, we propose new definitions of these terms and explore their consequences on the equilibrium structure of cooperative dilemmas in relation to social optimality. We find that a large class of multi-player prisoners' dilemmas and snowdrift games behave as their two-player counterparts, namely, they are characterized by a unique equilibrium where cooperation is always underprovided, regardless of the number of players. Multi-player stag hunts allow for the peculiarity of excessive cooperation at equilibrium, unless cooperation is such that it induces positive individual externalities. Our framework and results unify, simplify, and extend previous work on the structure and properties of binary-action multi-player cooperative dilemmas.

1 Introduction

Cooperative (or social) dilemmas can be informally described as situations where there is a tension between individual and collective interest regarding the cooperative behavior of individuals within a group (Dawes, 1980; Kollock, 1998; Hauert et al., 2006; Nowak, 2012; Rand and Nowak, 2013; Van Lange et al., 2013). The tension arises because cooperation can benefit the whole group but individuals might prefer to reduce their own cooperation and exploit the cooperative behavior of others. Examples of cooperative dilemmas include the private provision of public goods (Olson, 1965; Bergstrom et al., 1986), the management of common resources (Ostrom, 1990), voting (Palfrey and Rosenthal, 1983), protests, and other kinds of political

[*]Institute for Advanced Study in Toulouse, University of Toulouse Capitole, jorge.pena@iast.fr
[†]Institute for Advanced Study, University of Amsterdam
[‡]Faculty of Business and Economics, University of Basel, georg.noeldeke@unibas.ch

participation (Dawes et al., 1986), vaccination (Siegal et al., 2009), vigilance and sentinel behavior (Clutton-Brock et al., 1999), and many more.

Given their ubiquity, the study of cooperative dilemmas and their resolution has attracted enormous attention from a wide array of scholars in economics, political science, anthropology, psychology, evolutionary biology, and other disciplines. Across these different disciplines, game theory has emerged as the standard way of formalizing and thinking about cooperative dilemmas (Fudenberg and Tirole, 1991; Weibull, 1995; McNamara and Leimar, 2020). Within this perspective, a social interaction is conceptualized as a game whose equilibria predict the strategic behavior of individuals in the long run. Such equilibria are stable states expected to emerge as a result of individual rationality, individual or social learning, or of evolution acting on a population. The literature of cooperative dilemmas has used different equilibrium concepts, including the Nash equilibrium (NE), the evolutionarily stable strategy (ESS), and the asymptotic stable equilibrium (ASE) of the replicator dynamic (Taylor and Jonker, 1978). Here, we make use of the ESS as equilibrium concept and guiding principle. In simple terms, an ESS is a strategy such that if all members of a population adopt it, then no rare alternative strategy would fare better (Maynard Smith and Price, 1973). The ESS is an equilibrium refinement of the (symmetric) NE, and, for the games we consider in this paper, equivalent to the concept of ASE (Bukowski and Miekisz, 2004).

Conceivably, the simplest game-theoretic representation of a cooperative dilemma is as a symmetric game of complete information between players that can choose between two alternative actions or strategies ("cooperation" and "defection"), i.e., a multiplayer matrix game (Broom et al., 1997; Bukowski and Miekisz, 2004; Gokhale and Traulsen, 2014; Peña et al., 2014). The most paradigmatic example of such two-strategy cooperative dilemmas is the two-player prisoners' dilemma (see, e.g., Kollock, 1998). In this game, "defection" is a dominant strategy (so that it is individually optimal to defect regardless of the co-player's choice) and hence the only ESS (so that a population of defectors cannot be invaded by mutants cooperating with some probability). However, mutual "cooperation" yields higher payoffs to both players and can be, for certain payoff constellations, the socially optimal outcome. The (two-player) prisoners' dilemma is able to capture the essence of a cooperative dilemma in the starkest possible way, with a population trapped at an unique ESS featuring no cooperative behavior while expected payoffs would be maximized at some positive level of cooperation.

Although much earlier work focused exclusively on the prisoners' dilemma, it has been realized that in many situations two other two-player games can be better representations of cooperative dilemmas: the snowdrift (or chicken) game (Doebeli and Hauert, 2005), and the stag hunt (or assurance game) (Skyrms, 2004). While the prisoners' dilemma is characterized by both greed (an incentive to defect if the co-player cooperates) and fear (a disincentive to cooperate if the co-player defects), the snowdrift game is characterized by greed (but not fear) and the stag hunt is characterized by fear (but not greed). These different incentive structures lead to different ESS patterns. First, for the snowdrift game, there is a unique ESS characterized by a population where there is some cooperation, although less than what would maximize the expected payoff. Hence, in contrast to the prisoners' dilemma, some level of cooperation can be evolutionarily stable. However, as in the prisoners' dilemma, such level of cooperation is lower than the socially optimal level. Second, for the stag hunt, there are two ESSs: the first with full defection, and the second with full cooperation, and where the fully cooperative ESS coincides with the socially optimal level of cooperation. Hence, in contrast to the prisoners' dilemma, the socially optimal level of cooperation is evolutionarily stable. However, as in the prisoners' dilemma, the population can be trapped at the equilibrium where nobody cooperates. Taken together, the prisoners' dilemma, the snowdrift game, and the stag hunt constitute the three paradigmatic examples used to describe and think about cooperative dilemmas (Kollock, 1998).

In light of the wealth of research on cooperative and social dilemmas that has been published in recent decades, one would have anticipated a broad consensus regarding how to precisely define concepts such as "cooperation" and "cooperative dilemma", at the very least for symmetric matrix games. However, this does not appear to be the case. In fact, there are multiple coexisting definitions that are often at odds about the status of an action as cooperative (or not) or of a game as a cooperative dilemma (or not). Moving from two to more than two players only exacerbates the problem. Part of the issue is that many definitions proceed axiomatically by suggesting ways to classify games as cooperative dilemmas if given payoff inequalities hold, while other definitions emphasize the equilibrium structure (e.g., the ESS pattern) in relation to the location of socially optimal strategies that maximize expected payoffs. Such ambiguity is similar (and not unrelated) to the one surrounding the term "altruism" in evolutionary biology (Kerr et al., 2004).

Here, we build on previous work (Dawes, 1980; Kollock, 1998; Kerr et al., 2004; Peña et al., 2014, 2015) to propose definitions of "cooperation", "social dilemma", and "cooperative dilemma" that are internally consistent and that are useful to characterize the outcome of social interactions. We also propose multi-player generalizations of the trinity of games used in social dilemmas research, namely the prisoners' dilemma, the snowdrift game, and the stag hunt. We ask for these games if it is also the case, as it is for their well-known two-player counterparts, that cooperation is always underprovided at (an inefficient) equilibrium. A similar question has been asked before, although for more specific classes of cooperative dilemmas, by Gradstein and Nitzan (1990) and Anderson and Engers (2007).

The rest of this paper is organized as follows. We begin by presenting our general framework, and by establishing terminology, notation, and preliminary results in Section 2. Bernstein transforms, well known in approximation theory and computer-aided geometric design for decades and only more recently fully incorporated in game theory (Peña et al., 2014; Nöldeke and Peña, 2016), are important tools of our analysis. We then present our main definitions in Section 3. We define an action to be cooperative if two conditions hold (Definition 7). First, universal cooperation must provide higher payoffs than universal defection (Dawes, 1980). Second, cooperation must provide what we call "positive aggregate externalities", that is, a player switching from defection to cooperation must increase the aggregate payoff of co-players for any profile of pure strategies adopted by co-players (Matessi and Karlin, 1984; Kerr et al., 2004; Peña et al., 2015). Building on this definition, we then define a cooperative dilemma as a game with a cooperative action that is also a social dilemma (Definition 3). In turn, we define a social dilemma as a game featuring at least one ESS that is not socially optimal, in the sense that it does not maximize the expected payoff (Definition 2). This definition of social dilemma is in the spirit of the definition of the same term given by Kollock (1998), but adapted to our evolutionary (and symmetric) setup.

Section 4 deals with simpler conditions guaranteeing that a game is a cooperative dilemma. A necessary and sufficient condition is that individuals have, ex ante, individual incentives to defect (Proposition 1). Simpler necessary (but not sufficient) and sufficient (but not necessary) conditions are given in terms of the ex post individual incentives to defect, and hence in terms of simple inequalities involving the payoffs from the game. Section 5 provides similarly simple conditions for full cooperation to be socially optimal.

We propose definitions of prisoners' dilemmas, snowdrift games, and stag hunts for any number of players $n \geq 2$ in Section 6. In all cases, each such multi-player game has a cooperative action and an incentive structure that is reminiscent of its two-player counterpart. Prisoners' dilemmas are such that defection is (weakly) dominant. Individual incentives are thus characterized by both greed (of exploiting the cooperative behavior of others) and fear (of being exploited by the defective behavior of others). Snowdrift games are characterized by greed only, with incentives to defect if sufficiently many others cooperate. Stag hunts are characterized by fear only, with disincentives to cooperate if not enough others cooperate. In all cases, the ESS structure of these games is the same as their two-player versions. Our definitions for these three kinds of multi-player games rely on the ex post incentive structure and are thus stated in terms of inequalities at the level of payoffs of the game. We also introduce generalized prisoners' dilemmas, snowdrift games, and stag hunts (including the proper games as particular instances) that are defined in terms of their ex ante incentive structure.

We find that cooperation is underprovided at inefficient equilibria for all (generalized) prisoners' dilemmas and (generalized) snowdrift games—just as it is the case for the two-player versions of these games. Our finding extend previous results derived for specific cases of snowdrift games (Gradstein and Nitzan, 1990; Anderson and Engers, 2007) to the larger class of generalized snowdrift games. For (generalized) stag hunts, we find that it is possible to find cases with excessive cooperation, where the fully cooperative ESS supports more cooperation than what is socially optimal. However, this is the case only if the game does not feature what we call "positive individual externalities", that is, that a player switching from defection to cooperation increases the payoff of each co-player, for any symmetric profile of pure strategies adopted by co-players (Uyenoyama and Feldman, 1980; Kerr et al., 2004).

Finally, Section 7 offers some concluding remarks.

2 General framework

2.1 Multi-player symmetric two-strategy games

We consider a normal form game with two pure strategies (or actions, or choices) denoted by C and D. We focus on symmetric games among $n \geq 2$ players where all players assume the same role in the game, and where the payoff of any player depends only on its own choice and on the numbers of players choosing the two available actions. Throughout, "game" should be understood as "symmetric two-strategy game". We write P_k for the payoff of a player choosing C when k of their co-players choose C (and $n-1-k$ of their co-players choose D), and Q_k for the payoff of a player choosing D when k of their co-players choose C (and $n-1-k$ of their co-players choose D). Payoffs can be written in matrix form as

$$\begin{array}{c} \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cccccc} n-1 & \dots & k & \dots & 1 & 0 \end{array} \\ \begin{pmatrix} P_{n-1} & \dots & P_k & \dots & P_1 & P_0 \\ Q_{n-1} & \dots & Q_k & \dots & Q_1 & Q_0 \end{pmatrix} \end{array}. \tag{1}$$

We collect the parameters P_k and Q_k in the *payoff sequences* $\boldsymbol{P} = (P_0, P_1, \dots, P_{n-1}) \in \mathbb{R}^n$ and $\boldsymbol{Q} = (Q_0, Q_1, \dots, Q_{n-1}) \in \mathbb{R}^n$. We assume that $\boldsymbol{P} \neq \boldsymbol{Q}$ holds, so as to exclude the uninteresting case where payoffs are independent of the chosen actions. However, $P_k = Q_k$ may hold for some (but not all) values of $k = 0, 1, \dots, n-1$, so that games with non-generic payoffs are included in our framework. In a similar spirit, we assume that $\boldsymbol{P}$ and $\boldsymbol{Q}$ are not simultaneously constant, so as to exclude the uninteresting case where both payoff sequences are independent of k and hence of the actions chosen by co-players.

We denote by T_i the sum of payoffs to the n players when i players choose C and $n-i$ choose D. Such total payoffs are given by

$$T_i = iP_{i-1} + (n-i)Q_i, \; i = 0, 1, \dots, n. \tag{2}$$

We collect them in the *total payoff sequence* $\boldsymbol{T} = (T_0, T_1, \dots, T_n) \in \mathbb{R}^{n+1}$. The average payoff to the n players when i players choose C and $n-i$ choose D is then given by T_i/n. The *average payoff sequence*, collecting the average payoffs, is simply denoted by $\boldsymbol{T}/n$.

2.2 Private, external, and social gains

Suppose that out of the n players, k players play C and $n-k$ players play D. Fasten attention on one of the D-players and suppose that such a "focal player" switches its action from D to C while co-players keep their actions fixed, so that the focal player becomes the $(k+1)$-th C-player in the group (Kerr et al., 2004; Peña et al., 2015). As a result of this behavioral switch, the total payoff to the n players changes from T_k to T_{k+1}. We let

$$S_k = \Delta T_k = T_{k+1} - T_k, \ k = 0, 1, \ldots, n-1 \tag{3}$$

denote such a change in total payoffs, and call it the *social gain* induced by the focal player.

The social gain can be decomposed into two parts. First, as a result of the switch, the focal player experiences a change in payoff given by

$$G_k = P_k - Q_k, \ k = 0, \ldots, n-1. \tag{4}$$

We call this change in payoff the *private gain* enjoyed by the focal player.[1] Second, because of the focal's switch, each of its k co-players playing C experiences a change in payoff given by $\Delta P_{k-1} = P_k - P_{k-1}$, and each of its $n-1-k$ co-players playing D experiences a change in payoff given by $\Delta Q_k = Q_{k+1} - Q_k$. Overall, the focal's co-players experience an aggregate change in payoff given by

$$E_k = k\Delta P_{k-1} + (n-1-k)\Delta Q_k, \ k = 0, 1, \ldots, n-1, \tag{5}$$

where we set $P_{-1} = Q_n = 0$. We call this aggregate change the *external gain* or *aggregate externality* induced by the focal player.[2] Clearly, we have that

$$S_k = G_k + E_k, \ k = 0, 1, \ldots, n-1, \tag{6}$$

holds, so that the social gain is the sum of the private gain and the external gain. We collect the terms G_k in the *private gain sequence* $\boldsymbol{G} = (G_0, G_1, \ldots, G_{n-1}) \in \mathbb{R}^n$, the terms E_k in the *external gain sequence* or *aggregate externality sequence* $\boldsymbol{E} = (E_0, E_1, \ldots, E_{n-1}) \in \mathbb{R}^n$, and the terms S_k in the *social gain sequence* $\boldsymbol{S} = (S_0, S_1, \ldots, S_{n-1}) \in \mathbb{R}^n$.

The private gains $\boldsymbol{G}$ capture the individual incentives of a hypothetical focal player trying to determine his or her best choice given the choices of others. The choices of others are held fixed by fixing k, the number of co-players choosing C. A positive private gain ($G_k > 0$) then indicates an individual preference for choosing C over D (and hence an actual "gain" in payoff when hypothetically switching from D to C) while a negative private gain ($G_k < 0$) indicates an individual preference for choosing D over C (and hence a "loss" in payoff when switching from D to C). The private gains thus encapsulate the notions of *internality* suggested in Schelling (1973) and of *marginal private gain* discussed in Dixit et al. (2020, Ch. 11). The external gains $\boldsymbol{E}$, on the other hand, capture the spillover effects of the action of a focal player given the choices of co-players. A positive external gain ($E_k > 0$) indicates a positive spillover effect when choosing C over D (and hence an aggregate gain in payoff to co-players if the focal switches from D to C) while a negative external gain ($E_k < 0$) indicates a negative spillover effect when

[1] We have previously called such gain a "gain from switching" in Peña et al. (2014) and a "direct gain from switching" in Peña et al. (2015).

[2] We have previously called such gain an "indirect gain from switching" in Peña et al. (2015).

choosing C over D (and hence an aggregate loss in payoff to co-players in case the focal switches from D to C). The external gains encapsulate the notion of *marginal spillover effect* of Dixit et al. (2020, Ch. 11) and, more generally, of *externality*, which is "present whenever the behavior of a person affects the situation of other persons without the explicit agreement of that person or persons" (Buchanan, 1971, p. 7). The social gains $\boldsymbol{S}$ are the sum of private and external gains and thus capture the total effect of the switch of the focal and how it affects the total payoffs to the n players.

2.3 Sign patterns of sequences

To proceed, we need to specify how we will use the words positive, negative, increasing, and decreasing when referring to sequences, and to establish some terminology and notation to describe sign patterns of sequences (see, e.g., Brown et al. 1981; Peña et al. 2014). We need these definitions and terminology in order to capture in a precise way the qualitative features of ex post individual incentives ($\boldsymbol{G}$ sequence), externalities ($\boldsymbol{E}$ sequence), and social gains ($\boldsymbol{S}$ sequence) that characterize different kinds of games and cooperative dilemmas.

In the following, let $\boldsymbol{A} = (A_1, A_2, \ldots, A_m) \in \mathbb{R}^m$ be a non-zero vector (or sequence).

Positive and negative sequences. We say that $\boldsymbol{A}$ is non-negative, and write $\boldsymbol{A} \geq \mathbf{0}$, if $A_\ell \geq 0$ holds for all $\ell = 1, \ldots, m$. We say that $\boldsymbol{A}$ is positive, and write $\boldsymbol{A} \gneq \mathbf{0}$ if it is non-negative and non-zero, that is, if $A_\ell \geq 0$ holds for all $\ell = 1, \ldots, m$, with the inequality being strict for at least one ℓ. If the inequality is strict for all $\ell = 1, \ldots, m$ we say that $\boldsymbol{A}$ is strictly positive, and write $\boldsymbol{A} > \mathbf{0}$. Likewise, we say that $\boldsymbol{A}$ is non-positive, and write $\boldsymbol{A} \leq \mathbf{0}$, if $A_\ell \leq 0$ holds for all $\ell = 1, \ldots, m$. We say that $\boldsymbol{A}$ is negative, and write $\boldsymbol{A} \lneq \mathbf{0}$ if it is non-positive and non-zero. We say that it is strictly negative, and write $\boldsymbol{A} < \mathbf{0}$, if $A_\ell < 0$ holds for all $\ell = 1, \ldots, m$.

Increasing and decreasing sequences. Let us first define, for sequence $\boldsymbol{A}$, its first-forward difference $\Delta \boldsymbol{A} = (\Delta A_1, \ldots, \Delta A_{m-1}) \in \mathbb{R}^{m-1}$, where $\Delta A_\ell \equiv A_{\ell+1} - A_\ell$. We then say that $\boldsymbol{A}$ is increasing if $\Delta \boldsymbol{A}$ is positive, and that it is is non-increasing if $\Delta \boldsymbol{A}$ is non-positive; i.e., a non-increasing sequence is either constant or decreasing. Likewise, we say that $\boldsymbol{A}$ is decreasing if $\Delta \boldsymbol{A}$ is negative, and that it is is non-decreasing if $\Delta \boldsymbol{A}$ is non-negative; i.e., a non-decreasing sequence is either constant or increasing.

Sign changes, initial and final sign. We denote by $\sigma(\boldsymbol{A})$ the number of sign changes of $\boldsymbol{A}$, *ignoring zeros*. We also call the sign of the first non-zero element A_f of $\boldsymbol{A}$ the *initial sign* of $\boldsymbol{A}$, denote it by $\iota(\boldsymbol{A})$, and write $\iota(\boldsymbol{A}) = \operatorname{sgn}(a_f)$, where

$$\operatorname{sgn} x = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases} \tag{7}$$

is the sign function. Thus, $\iota(\boldsymbol{A}) = 1$ if the initial sign is positive and $\iota(\boldsymbol{A}) = -1$ if the initial sign is negative. Likewise, we call the sign of the last non-zero element of $\boldsymbol{A}$ the *final sign* of $\boldsymbol{A}$, denote it by $\phi(\boldsymbol{A})$, and write $\phi(\boldsymbol{A}) = 1$ if it is positive, and $\phi(\boldsymbol{A}) = -1$ if it is negative.

Sign pattern. Finally, we denote by $\varrho(\boldsymbol{A})$ the *sign pattern* of $\boldsymbol{A}$ the vector $\varrho(\boldsymbol{A}) \in \mathbb{R}^{\sigma(\boldsymbol{A})+1}$ obtained by (i) applying the sign function (7) element-wise to the vector $\boldsymbol{A}$, and (ii) removing zeros and consecutive repeated values. If $\boldsymbol{A}$ is positive (resp. negative) then, clearly, $\varrho(\boldsymbol{A}) = (1)$ (resp. $\varrho(\boldsymbol{A}) = (-1)$).

As an example to illustrate these definitions consider the sequence $\boldsymbol{A} = (0, 0, 1, 2, -3, 0, 4, -5)$. Then $\iota(\boldsymbol{A}) = 1$, $\phi(\boldsymbol{A}) = -1$, $\sigma(\boldsymbol{A}) = 3$, and $\varrho(\boldsymbol{A}) = (1, -1, 1, -1)$.

2.4 A bestiary of games

We illustrate the meaning of the different sequences that we have introduced so far, and our sign pattern terminology with the following examples of (multi-player) games. In the rest of the paper, we will repeatedly come back to these examples to illustrate our general framework, our definitions, and our main results.

Example 1 (Two-player games)**.** For $n = 2$ players, the payoff matrix (1) reduces to

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} P_1 & P_0 \\ Q_1 & Q_0 \end{pmatrix} \end{array}. \tag{8}$$

The payoff sequences are then $\boldsymbol{P} = (P_0, P_1)$, and $\boldsymbol{Q} = (Q_0, Q_1)$, the total payoff sequence is $\boldsymbol{T} = (2Q_0, P_0 + Q_1, 2P_1)$, the gain sequence is $\boldsymbol{G} = (P_0 - Q_0, P_1 - Q_1)$, the aggregate externality sequence is $\boldsymbol{E} = (Q_1 - Q_0, P_1 - P_0)$, and the social gain sequence is $\boldsymbol{S} = (P_0 + Q_1 - 2Q_0, 2P_1 - P_0 - Q_1)$.

Example 2 (Two-player cooperative dilemmas with generic payoffs)**.** Consider the three types of two-player cooperative dilemmas typically distinguished in the literature: the prisoners' dilemma, the snowdrift game, and the stag hunt. Assuming that payoffs are generic (i.e., no two payoff values are equal to each other), each of these two-player games is characterized by a particular ordering of the values of payoff matrix (8):

1. If $Q_1 > P_1 > Q_0 > P_0$, then the game is a prisoners' dilemma. In this case, the gain sequence is strictly negative (i.e., $\boldsymbol{G} < \boldsymbol{0}$) and hence $\sigma(\boldsymbol{G}) = 0$, $\iota(\boldsymbol{G}) = \phi(\boldsymbol{G}) = -1$, and $\varrho(\boldsymbol{G}) = (-1)$ hold.

    (a) If $2P_1 \geq P_0 + Q_1$, then the social gain sequence has sign pattern $\varrho(\boldsymbol{S}) = (1)$ if $P_0 + Q_1 \geq 2Q_0$ holds and $\varrho(\boldsymbol{S}) = (-1, 1)$ otherwise.

(b) If $2P_1 < P_0 + Q_1$, then the social gain sequence has sign pattern $\varrho(\boldsymbol{S}) = (1, -1)$.

2. If $Q_1 > P_1 > P_0 > Q_0$, then the game is a snowdrift game (or a chicken game). In this case, the gain sequence has a single sign change from positive to negative so that $\sigma(\boldsymbol{G}) = 1$, $\iota(\boldsymbol{G}) = 1$, $\phi(\boldsymbol{G}) = -1$, and $\varrho(\boldsymbol{G}) = (1, -1)$ hold.

   (a) If $2P_1 \geq P_0 + Q_1$, then the social gain sequence has sign pattern $\varrho(\boldsymbol{S}) = (1)$.

   (b) If $2P_1 < P_0 + Q_1$, then the social gain sequence has sign pattern $\varrho(\boldsymbol{S}) = (1, -1)$.

3. If $P_1 > Q_1 > Q_0 > P_0$, then the game is a stag hunt (or assurance game). In this case, the gain sequence has a single sign change from negative to positive, and so $\sigma(\boldsymbol{G}) = 1$, $\iota(\boldsymbol{G}) = -1$, $\phi(\boldsymbol{G}) = 1$, and $\varrho(\boldsymbol{G}) = (-1, 1)$ hold. As the inequality $2P_1 > P_0 + Q_1$ holds, the social gain sequence has sign pattern $\varrho(\boldsymbol{S}) = (1)$ or $\varrho(\boldsymbol{S}) = (-1, 1)$.

In all three cases, the external gain sequence $\boldsymbol{E}$ is strictly positive, i.e., $\boldsymbol{E} > \boldsymbol{0}$ holds.

Example 3 (Public goods games (general))**.** Consider public goods games where playing C means to voluntarily contribute to a public good while playing D means to shirk (as considered by, e.g., Taylor and Ward, 1982; Rapoport, 1987; Gradstein and Nitzan, 1990; Weesie and Franzen, 1998; Dixit and Olson, 2000; Hauert et al., 2006; Makris, 2009; Pacheco et al., 2009; Souza et al., 2009; Archetti and Scheuring, 2011; Santos and Pacheco, 2011; Peña et al., 2014; De Jaegher, 2017). Contributing entails a cost $c_i \geq 0$ to each C-player, while all players (both C-players and D-players) enjoy a benefit $b_i \geq 0$, where $0 \leq i \leq n$ denotes the total number of players choosing C. The payoff sequences $\boldsymbol{P}$ and $\boldsymbol{Q}$ are then given by

$$P_k = b_{k+1} - c_{k+1},\ k = 0, 1, \ldots, n-1 \tag{9}$$

$$Q_k = b_k,\ k = 0, 1, \ldots, n-1. \tag{10}$$

We collect the costs in the cost sequence $\boldsymbol{c} = (c_0, c_1, \ldots, c_n) \in \mathbb{R}^{n+1}$ and the benefits in the benefit sequence $\boldsymbol{b} = (b_0, b_1, \ldots, b_n) \in \mathbb{R}^{n+1}$. We assume that $\boldsymbol{b}$ is increasing (so that the larger the number of C-players, the larger the value of the public good that is provided) and that $\boldsymbol{c}$ is non-decreasing (so that increasing the number of C-players never increases the cost associated to contributing). We further assume that $b_{n-1} - b_0 > c_n$ holds (i.e., that the difference between the value of the public good when everybody contributes and its value when nobody contributes is larger than the personal cost when everybody contributes).

Since benefits $\boldsymbol{b}$ are increasing and costs $\boldsymbol{c}$ are non-decreasing, the payoff sequences $\boldsymbol{P}$ and $\boldsymbol{Q}$ are increasing: Every player is better off the more other players contribute to the public good. It follows from Eq. (5) that the aggregate externality sequence is positive ($\boldsymbol{E} \gneqq \boldsymbol{0}$), i.e., contributing to the public good has positive spillover effects. This can also be verified by inspection of the external gains, which are given by

$$E_k = (n-1)\Delta b_k - k\Delta c_k,\ k = 0, 1, \ldots, n-1. \tag{11}$$

The total payoffs are given by

$$T_i = nb_i - ic_i, \ i = 0, 1, \ldots, n, \tag{12}$$

i.e., by the difference between the total benefits (nb_i) and the total costs (ic_i) in a group of n players, i of which contribute to the collective action. The social gains thus satisfy

$$S_k = \Delta T_k = n\Delta b_k - [(k+1)c_{k+1} - kc_k], \ k = 0, 1, \ldots, n-1. \tag{13}$$

Since $\boldsymbol{b}$ is increasing, a sufficient condition for $\boldsymbol{S}$ to be positive (and $\boldsymbol{T}$ to be increasing) is then that

$$(k+1)c_{k+1} \leq kc_k, \ k = 0, 1, \ldots, n-1 \tag{14}$$

holds, i.e., that the total costs borne by contributors is non-increasing in the number of contributors.

The private gains are given by

$$G_k = \Delta b_k - c_{k+1}, \ k = 0, 1, \ldots, n-1. \tag{15}$$

The sign pattern of the private gain sequence $\boldsymbol{G}$ depends on the particular shapes of the benefit and the cost sequences, and in particular on how the marginal benefit contributing $\Delta\boldsymbol{b}$ scales with the number of contributors and compares to the cost $\boldsymbol{c}$. Particular examples are given in Examples 4, 5, 6, and 7 below.

Example 4 (Public goods games with concave benefits and fixed costs)**.** Consider a particular instance of the public goods game defined in Example 3 where $\boldsymbol{b}$ is concave (i.e., $\Delta^2\boldsymbol{b}$ is negative) and $\boldsymbol{c}$ is constant of value $\gamma > 0$ (i.e., $\boldsymbol{c} = (\gamma, \gamma, \ldots, \gamma)$), as assumed, e.g., by Gradstein and Nitzan (1990) and Motro (1991).[3] Then $\Delta\boldsymbol{G} \lneq \boldsymbol{0}$ holds and the private gain sequence $\boldsymbol{G}$ is decreasing. If costs are high ($\gamma \geq \Delta b_0$), $\boldsymbol{G}$ is negative. If costs are low ($\gamma \leq \Delta b_{n-1}$), $\boldsymbol{G}$ is positive. If costs are intermediate (i.e., $\Delta b_{n-1} < \gamma < \Delta b_0$ holds), $\boldsymbol{G}$ has a single sign change from positive to negative, and the sign pattern of $\boldsymbol{G}$ is $\varrho(\boldsymbol{G}) = (1, -1)$. In this case, players have an individual incentive to contribute to the public good when there are relatively few contributors, and they have an incentive to shirk when there are relatively many contributors.

Example 5 (Public goods games with convex benefits)**.** As a second subclass of the general public goods game introduced in Example 3, suppose that $\boldsymbol{b}$ is convex (i.e., $\Delta^2\boldsymbol{b}$ is positive). Then, without the need of further assumptions on the cost sequence, $\Delta\boldsymbol{G} \gneq \boldsymbol{0}$ holds and the private gain sequence $\boldsymbol{G}$ is increasing. If costs are high ($c_n \geq \Delta b_{n-1}$), $\boldsymbol{G}$ is negative. If costs are low ($c_1 \leq \Delta b_0$), $\boldsymbol{G}$ is positive. If costs are intermediate (i.e., $\Delta b_0 < c_1$ and $\Delta b_{n-1} > c_n$ hold), $\boldsymbol{G}$ has a single sign change from negative to positive, so that the sign pattern of $\boldsymbol{G}$ is

[3] They assume strict concavity of $\boldsymbol{b}$, i.e., $\Delta^2\boldsymbol{b} > \boldsymbol{0}$. Our condition is more relaxed.

$\varrho(\boldsymbol{G}) = (-1, 1)$.

Example 6 (Public goods games with sigmoid benefits and fixed costs)**.** As a third subclass of the general public goods game introduced in Example 3, suppose that $\boldsymbol{b}$ is first convex, then concave (i.e., $\Delta^2 \boldsymbol{b}$ has a single sign change from positive to negative), and $\boldsymbol{c}$ is constant of value $\gamma > 0$ (i.e., $\boldsymbol{c} = (\gamma, \gamma, \ldots, \gamma)$). Examples include models discussed by Pacheco et al. (2009) and Archetti and Scheuring (2011), where the benefit sequence first accelerates and then decelerates with the number of contributors. Since $\Delta \boldsymbol{G} = \Delta^2 \boldsymbol{b}$ holds, it follows that $\Delta \boldsymbol{G}$ has a single sign change from positive to negative, which means that the private gain function is unimodal, i.e., first increasing, then decreasing. Then, depending on how the cost of contributing γ relates to $\Delta \boldsymbol{b}$, we have the following cases. If costs are high ($\gamma \geq \max_k \Delta b_k$), $\boldsymbol{G}$ is negative. If costs are low ($\gamma \leq \min_k \Delta b_k$), $\boldsymbol{G}$ is positive. If costs are intermediate (i.e., $\min_k \Delta b_k < \gamma < \max_k \Delta b_k$ holds), then the sign pattern of the private gain sequence $\varrho(\boldsymbol{G})$ depends on the relative position of Δb_0 and Δb_{n-1} with respect to γ, as follows. If $\Delta b_0 \geq \gamma$ and $\Delta b_{n-1} < \gamma$, then $\varrho(\boldsymbol{G}) = (1, -1)$, just as in Example 4. If $\Delta b_0 < \gamma$ and $\Delta b_{n-1} \geq \gamma$, then $\varrho(\boldsymbol{G}) = (-1, 1)$, just as in Example 5. Finally, if $\max\{\Delta b_0, \Delta b_{n-1}\} < \gamma$, then $\varrho(\boldsymbol{G}) = (-1, 1, -1)$. This case, where the private gain sequence has two sign changes (the first one from negative to positive, the second from positive to negative) is different from the previous examples. Here, players have an incentive to contribute to the public good only if sufficiently many (but not too many) other players also contribute.

Example 7 (Threshold public goods game with fixed costs)**.** A noteworthy example of a public goods game is the threshold public goods game with fixed costs and no refunds (Taylor and Ward, 1982; Palfrey and Rosenthal, 1984; Bach et al., 2006; Nöldeke and Peña, 2020). In this game, contributors pay a non-refundable cost equal to $0 < \gamma < 1$ and the public good is provided if and only if the number of contributors reaches an exogenous threshold θ, in which case all players get the same benefit (normalized to one) from the provision of the public good. The cost sequence is thus given by $\boldsymbol{c} = (\gamma, \gamma, \ldots, \gamma)$ and the benefit sequence by

$$b_i = [\![i \geq \theta]\!], \; i = 0, 1, \ldots, n, \tag{16}$$

where $[\![\,]\!]$ denotes the Iverson bracket, i.e., $[\![X]\!] = 1$ if X is true and $[\![X]\!] = 0$ if X is false. If $\theta = 1$ (only one contributor is required) the game is known as the "volunteer's dilemma" (Diekmann, 1985). In this case, $\boldsymbol{b}$ is concave, and $\Delta b_{n-1} = 0 < \gamma < 1 = \Delta b_0$ holds. Hence, in this case the game is a particular instance of the subclass of public goods games with concave benefits and fixed intermediate costs presented in Example 4. In particular, the sign pattern of the private gain sequence is $\varrho(\boldsymbol{G}) = (1, -1)$. Alternatively, if $\theta = n$ (all contributors are required), $\boldsymbol{b}$ is convex, and both $\Delta b_0 = 0 < \gamma = c_1$ and $\Delta b_{n-1} = 1 > \gamma = c_n$ hold. Hence, in this case the game is a particular instance of the subclass of public goods games with convex benefits and intermediate costs presented in Example 5. In particular, the sign pattern of the private gain sequence is $\varrho(\boldsymbol{G}) = (-1, 1)$. Finally, if $1 < \theta < n$ holds (more than one but less than all contributors are needed) the game is sometimes referred to as the "teamwork dilemma" (Myatt and Wallace, 2008; Nöldeke and Peña, 2020). In this case, the gain sequence is given by $G_k = -\gamma < 0$ for $k \neq \theta - 1$ and $G_{\theta-1} = 1 - \gamma > 0$, and hence $\varrho(\boldsymbol{G}) = (-1, 1, -1)$ holds. Here, individuals have an incentive to contribute to the public good if and only if exactly other $\theta - 1$ players were to contribute, as only in such scenario their contribution is required (or pivotal). The "teamwork dilemma" is a particular case of the subclass of public goods games with sigmoid benefits and fixed costs presented in Example 6.

Example 8 (Participation games with negative externalities (congestion games))**.** Consider the class of participation games with negative externalities to other participants (or congestion games) discussed in Anderson and Engers (2007, Section 3). This class includes, for instance, the threshold participation game with "negative feedback" of Dindo and Tuinstra (2011) and Arthur (1994)'s El Farol bar problem. Playing D (to participate, or to choose "in") means to take part in an activity such as entering a market, exploiting a common resource, driving, or going to a bar. Playing C (to abstain from participating, or to stay "out") means to refrain from taking part in such an activity. The payoff to choosing "out" is a constant $\gamma > 0$ (that Anderson and Engers (2007) normalize to zero). The payoff to choosing "in" is a decreasing function of the total number of D-players. Thus, participants generate negative externalities to other participants. The payoff sequences $\boldsymbol{P}$ and $\boldsymbol{Q}$ are given by

$$P_k = \gamma, \; k = 0, 1, \ldots, n-1 \tag{17}$$

$$Q_k = v_{n-1-k}, \; k = 0, 1, \ldots, n-1, \tag{18}$$

where v_ℓ, $\ell = 0, 1, \ldots, n-1$ is the value of the activity to a participant (D-player) given the number ℓ of other participants among co-players. By assumption, the sequence $\boldsymbol{v} = (v_0, v_1, \ldots, v_{n-1}) \in \mathbb{R}^n$ is decreasing, i.e., $\Delta \boldsymbol{v} \lneq \boldsymbol{0}$ holds. It follows that $\boldsymbol{P}$ is constant and $\boldsymbol{Q}$ is increasing, i.e., $\Delta \boldsymbol{P} = \boldsymbol{0}$ and $\Delta \boldsymbol{Q} \gneq \boldsymbol{0}$ hold.

The private gains are given by

$$G_k = \gamma - v_{n-1-k}, \; k = 0, 1, \ldots, n-1. \tag{19}$$

It is assumed that $v_0 > \gamma > v_{n-1}$ holds, so that the payoff to play "in" when everybody else plays "out" is greater than the payoff to play "out", which is in turn greater than the payoff to play "in" when everybody else plays "in". The private gain sequence $\boldsymbol{G}$ is thus decreasing (i.e., $\Delta \boldsymbol{G} \lneq \boldsymbol{0}$) and characterized by the sign pattern $\varrho(\boldsymbol{G}) = (1, -1)$. That is, players have an incentive to participate in the activity (entering a market, exploiting a common resource, driving, going to a bar) as long as not too many others also decide to do so.

The external gains are given by

$$E_k = (n - 1 - k)\Delta v_{n-1-k}, \; k = 0, 1, \ldots, n-1, \tag{20}$$

which are always non-negative and sometimes positive. The external gain sequence $\boldsymbol{E}$ is thus positive (i.e., $\boldsymbol{E} \gneq \mathbf{0}$). In other words, not participating generates positive externalities to the aggregate of co-players.

Example 9 (Games with participation synergies (strategic complements in participation))**.** Consider the class of participation games with positive externalities to other participants discussed in Anderson and Engers (2007, Section 4), which are the counterpart to the class of games discussed in Example 8, and include the "club goods" studied in Peña et al. (2015) and the "n-person stag hunt game" of Luo et al. (2021). Let us now label C the decision to participate, or to choose "in", and D the decision to abstain from participating, or staying "out". As for congestion games, the payoff to staying "out" is a constant $\gamma > 0$ (that Anderson and Engers (2007) normalize to zero). The payoff to choosing "in" is now increasing in the number of other C-players. Thus, participants generate positive externalities to other participants. The payoff sequences $\boldsymbol{P}$ and $\boldsymbol{Q}$ are given by

$$P_k = v_{k+1}, \ k = 0, 1, \ldots, n-1 \tag{21}$$

$$Q_k = \gamma, \ k = 0, 1, \ldots, n-1, \tag{22}$$

where v_i, $i = 0, 1, \ldots, n$ is the value of the activity to a participant (C-player) given the total number i of participants among players (including the self). By assumption, the sequence $\boldsymbol{v} = (v_0, v_1, \ldots, v_n) \in \mathbb{R}^{n+1}$ is increasing, i.e., $\Delta \boldsymbol{v} \gneq \mathbf{0}$ holds. It follows that $\boldsymbol{P}$ is increasing and $\boldsymbol{Q}$ is constant. Luo et al. (2021) considered a particular case with $v_i = \beta [\![i \geq \theta]\!]$, $1 < \theta \leq n$, and $\beta > \gamma$.

The private gains are given by

$$G_k = v_{k+1} - \gamma, \ k = 0, 1, \ldots, n-1. \tag{23}$$

Since $\boldsymbol{v}$ is increasing, so is the private gain sequence $\boldsymbol{G}$. It is also assumed that $v_0 < \gamma < v_n$ holds, so that the payoff to play "in" when everybody else plays "out" is smaller than the payoff to play "out", which is in turn smaller than the payoff to play "in" when everybody else plays "in". Hence, the private gain sequence has sign pattern $\varrho(\boldsymbol{G}) = (-1, 1)$. In this case, players have an incentive to participate in the activity as long as sufficiently many others also decide to do so.

The external gains are given by

$$E_k = k \Delta v_k, \ k = 0, 1, \ldots, n-1, \tag{24}$$

so that the external gain sequence $\boldsymbol{E}$ is positive (i.e., $\boldsymbol{E} \gneq \mathbf{0}$). Here, participating generates positive externalities to the aggregate of co-players.

2.5 Mixed strategies and expected payoffs

We consider mixed strategies represented by $\boldsymbol{x} \in \Delta^1 \equiv \{(x, 1-x) \mid 0 \leq x \leq 1\}$, where Δ^1 is the 1-simplex. Pure strategy C (resp. D) corresponds to mixed strategy $\boldsymbol{x} = (1, 0)$ (resp. $\boldsymbol{x} = (0, 1)$). We call mixed strategies $\boldsymbol{x} = (x, 1-x)$ with $x \in (0, 1)$, *totally mixed strategies*. There are two alternative interpretations of a mixed strategy $\boldsymbol{x} = (x, 1-x)$. The first, common in classic game theory and static evolutionary game theory (and relevant for the notions of symmetric NE and ESS), is that the mixed strategy represents the strategy played by a given player (or the phenotype of a given individual). In this case, x (resp. $1-x$) represents the probability that this player chooses action C (resp. D). The second interpretation of a mixed strategy, common in dynamic evolutionary game theory (and relevant for the notion of an ASE), is as a population state in a large population of players using pure strategies. In this case, x corresponds to the proportion of individuals in the population using pure strategy C (or C-players), and $1-x$ corresponds to the proportion of individuals using pure strategy D (or D-players). Both interpretations have been used in the literature of two-strategy cooperative dilemmas, and we find it useful to have them both in mind. In the following, we refer to mixed strategy $\boldsymbol{x} = (x, 1-x)$ simply by x. For our analysis, it suffices to focus on symmetric profiles where all co-players of a given player play the same mixed strategy x.

Let us adopt here the first interpretation of a mixed strategy. Writing $f_C(x)$ (resp. $f_D(x)$) for the expected payoff to a C-player (resp. D-player) when all co-players play x, we have

$$f_C(x) = \sum_{k=0}^{n-1} \binom{n-1}{k} x^k (1-x)^{n-1-k} P_k, \tag{25a}$$

$$f_D(x) = \sum_{k=0}^{n-1} \binom{n-1}{k} x^k (1-x)^{n-1-k} Q_k. \tag{25b}$$

Indeed, with the binomial probability $\binom{n-1}{k} x^k (1-x)^{n-1-k}$ a focal player choosing C (resp. D) will have k co-players having chosen C, in which case he or she will obtain a payoff of P_k (resp. Q_k). Summing over all possibilities weighted by the given payoff, we obtain the expected payoff to the focal player.

We are interested in the *expected payoff* of a player playing mixed strategy x when all co-players also play x. This is because, as it will be explained below, any symmetric profile x that does not maximize the expected payoff will be regarded as inefficient, while the symmetric profile x that maximizes the expected payoff will be regarded as socially optimal. The *expected payoff*, which we denote by $f(x)$, is given by

$$f(x) = x f_C(x) + (1-x) f_D(x). \tag{26}$$

Indeed, a focal player playing strategy x will choose action C with probability x, in which case its expected payoff when co-players also play x is $\pi_C(x)$, and with probability $1-x$ it will choose action D, in which case its expected payoff is $\pi_D(x)$. Substituting from Eqs. (25) this can be alternatively written as

$$f(x) = \sum_{i=0}^{n} \binom{n}{i} x^i (1-x)^{n-i} \frac{T_i}{n}, \tag{27}$$

i.e., as the expected average payoff to a group of n players playing x.

2.6 Private, external, and social gain functions

The marginal change in expected payoff when players change their mixed strategy infinitesimally is given by the derivative f' of the expected payoff function f. We call this derivative the *social gain function*. By differentiating (27) and simplifying, the social gain function can be written as

$$f'(x) = \sum_{k=0}^{n-1} \binom{n-1}{k} x^k (1-x)^{n-1-k} S_k. \tag{28}$$

Note that this is nothing but the expected social gains when the choice of one focal player is changed from C to D and the number of co-players choosing C among the co-players of a focal player is distributed according to a binomial distribution with parameters $n-1$ and x.

Clearly, since the social gains equal the private gains plus the external gains (see Eq. (6)), we obtain after rearranging

$$f'(x) = g(x) + h(x), \tag{29}$$

where

$$g(x) = \sum_{k=0}^{n-1} \binom{n-1}{k} x^k (1-x)^{n-1-k} G_k, \tag{30}$$

is the *private gain function*, and

$$h(x) = \sum_{k=0}^{n-1} \binom{n-1}{k} x^k (1-x)^{n-1-k} E_k, \tag{31}$$

is the *external gain function* or the *aggregate externality function*.

The private gain function (30) corresponds to the expected private gains (4) induced by a focal player switching its action from D to C when the number of co-players choosing C is distributed binomially with parameters $n-1$ and x. The private gain function thus tells us by how much a switch from D to C increases the focal's expected payoff when all $n-1$ co-players of a focal player randomize their actions independently with probability x of choosing C. Similarly, the external gain function (31) corresponds to the expected external gain (5) induced by the focal's switch and tells us how much the aggregate expected payoff of co-players changes in such a situation because of the focal's switch. Eq. (29) is then the statement that the effect on the expected average payoff of a marginal increase in x for all players is given by the sum of the private gain and external gain resulting from switching the action of one player while keeping the mixed strategy of all other players fixed at x.

2.7 Sign patterns of polynomials

Before proceeding, and similarly to the way we did for sequences in Section 2.3, we need to specify some terminology and notation referring to the properties of polynomials like the three gain functions we introduced in the preceding section. We need these definitions and terminology in order to capture in a precise way the qualitative features of ex ante individual incentives (g function), externalities (h function), and social gains (f' function) that characterize different kinds of games and cooperative dilemmas.

In the following, consider a polynomial $p : [0, 1] \to \mathbb{R}$.

Positive and negative polynomials. We will say that p is positive, and write $p \gneq 0$ if $p(x) \geq 0$ holds for all $x \in [0, 1]$ and the inequality is strict for at least some $x \in (0, 1)$. We say that p is strictly positive, and write $p > 0$ if $p(x) > 0$ holds for all $x \in (0, 1)$. Likewise, we say that p is negative, and write $p \lneq 0$ if $p(x) \leq 0$ holds for all $x \in [0, 1]$ and the inequality is strict for at least some $x \in (0, 1)$. We say that p is strictly negative, and write $p < 0$ if $p(x) < 0$ holds for all $x \in (0, 1)$.

Increasing and decreasing polynomials. Let us denote by p' the derivative of polynomial p. We then say that p is increasing if p' is positive, and that it is is non-increasing if p' is non-positive; i.e., a non-increasing polynomial is either constant or decreasing. Likewise, we say that p is decreasing if p' is negative, and that it is is non-decreasing if p' is non-negative; i.e., a non-decreasing polynomial is either constant or increasing.

Sign changes. We say that p changes sign from positive to negative (resp. negative to positive) at a point $x \in (0, 1)$ if (i) $p(x) = 0$ and, for y close to x, both of these two implications hold: (iia) if $y < x$ then $p(y) > 0$ (resp. $p(y) < 0$), and (iib) if $y > x$ then $p(y) < 0$ (resp. $p(y) > 0$). In general, we say that p changes sign at a point $x \in (0, 1)$ if it changes sign from positive to negative or from negative to positive.

Number of sign changes. We denote by $\sigma(p)$ the number of sign changes of p. The number of sign changes $\sigma(p)$ is equal to the number of times p crosses the x-axis in $(0, 1)$.

Initial and final signs. Assume $p \neq 0$ holds. Then there exists a neighborhood of $x = 0$ such that the sign of p is either positive or negative throughout this neighborhood. We then define the *initial sign* of p as the sign of p in such neighborhood, denote it by $\iota(p)$, and write $\iota(p) = 1$ if it is positive, and $\iota(p) = -1$ if it is negative. Similarly, there exists a neighborhood of $x = 1$ such that the sign of p is either positive or negative throughout this neighborhood and we can define the final sign of p as $\phi(p) = 1$ if p is positive in such a neighborhood and $\phi(p) = -1$ if it is negative. Clearly, $\iota(p) = \text{sgn}(p(0))$ if $p(0) \neq 0$ holds. Similarly, $\phi(f) = \text{sgn}(p(1))$ if $p(1) \neq 0$ holds.

Sign pattern. The *sign pattern* of p is given by a sequence $\varrho(p) \in \mathbb{R}^{\sigma(p)+1}$ with alternating ones and minus ones with its first element given by $\iota(p)$. The sign pattern describes the sign variations of the polynomial p, conveniently summarizing all the information on initial signs, final signs, and sign changes.

2.8 Private gain function and evolutionary stability

The private gain function guides both individual behavior (in a non-evolutionary context of maximizing rational agents) and individual selection (in an evolutionary context under a simple demography where kin selection or group selection do not play any role). Indeed, which mixed strategies turn to be symmetric NE, ESS, or ASE is fully determined by the private gain function.[4]

For our subsequent analysis and in the rest of this paper, we use the ESS as our solution concept. The notion of ESS is a refinement of symmetric NE, as every ESS is a symmetric NE, but the converse is not true (Bukowski and Miekisz, 2004, Theorem 6). Also, for two-strategy symmetric n-player games, as the ones we focus on, the notions of ESS and ASE imply each other, i.e., every ESS is an ASE and every ASE is an ESS (Bukowski and Miekisz, 2004, Corollary 2). It follows that we can restate all of our results in terms of the stable rest points of the replicator dynamic, instead of the corresponding evolutionarily stable strategies.

In the following, we present simple conditions for a mixed strategy x to be an ESS. Since we have assumed that $\boldsymbol{P} \neq \boldsymbol{Q}$ holds, $\boldsymbol{G} \neq \boldsymbol{0}$ holds. From Eq. (30) this in turn implies $g \neq 0$, so that the initial sign $\iota(g)$ and final sign $\phi(g)$ of g are well defined. We can then state the following result, which is simply a restatement of Bukowski and Miekisz 2004, Theorem 3:

Lemma 1 (Sign pattern of g and evolutionary stability)**.** Let g be the gain function of a symmetric two-strategy n-player game, with initial sign $\iota(g)$ and final sign $\phi(g)$. Then

1. $x^* = 0$ is an ESS if and only if the initial sign of g is negative, i.e., $\iota(g) = -1$.

2. $x^* = 1$ is an ESS if and only if the final sign of g is positive, i.e., $\phi(g) = 1$.

3. $x^* \in (0, 1)$ is an ESS if and only if g changes sign from positive to negative at x^*.

Lemma 1 provides a convenient link between the sign pattern of the private gain function, and the ESS structure of the underlying multi-player game.

[4] Regarding necessary and sufficient conditions for a symmetric NE, we have: (i) $x = 0$ is a symmetric NE if and only if $g(0) \leq 0$, (ii) $x = 1$ is a symmetric NE if and only if $g(1) \geq 0$, and (iii) $x \in (0, 1)$ is a symmetric NE if and only if $g(x) = 0$, i.e., if it is a root of g.

2.9 Social gain function and social optimality

In addition to evolutionarily stable strategies, an important concept in our analysis is the *social optimum*, which we define as the mixed strategy

$$\hat{x} = \arg\max_{x \in [0,1]} f(x) \tag{32}$$

that maximizes the expected payoff (27). For simplicity, we assume that such an optimum is unique.

In our general framework, the social optimum corresponds to either one of the two pure strategies (i.e., $\hat{x} = 0$ or $\hat{x} = 1$) or to a totally mixed strategy $\hat{x} \in (0, 1)$. Since the social gain function f' is the derivative of the expected payoff f, the sign pattern of the social gain function f' provides necessary conditions for a mixed strategy to be a social optimum. In particular, a global maximum must be a local maximum. This observation leads to the following characterization, which links the sign pattern of f' to social optimality in a similar way Lemma 1 links the sign pattern of g to evolutionary stability.[5]

Lemma 2 (Sign pattern of f' and social optimality)**.** Let f' be the social gain function of a symmetric two-strategy n-player game, with initial sign $\iota(f')$ and final sign $\phi(f')$. Then

1. If $\hat{x} = 0$ is a social optimum then the initial sign of f' is negative, i.e., $\iota(f') = -1$.
2. If $\hat{x} = 1$ is a social optimum then the final sign of f' is positive, i.e., $\phi(f') = 1$.
3. If $\hat{x} \in (0, 1)$ is a social optimum then f' changes sign from positive to negative at $\hat{x}$.

2.10 The expected payoff and the gain functions are polynomials in Bernstein form

The expression

$$p(x) = \sum_{k=0}^{m} \binom{m}{k} x^k (1-x)^{m-k} c_k \equiv \mathcal{B}_m(x; \boldsymbol{c}) \tag{33}$$

is a *polynomial in Bernstein form*, i.e., a linear combination of the Bernstein basis polynomials

$$\binom{m}{k} x^k (1-x)^{m-k}, \; k = 0, 1, \ldots, m, \tag{34}$$

with coefficients given by the sequence $\boldsymbol{c} = (c_0, c_1, \ldots, c_m) \in \mathbb{R}^{m+1}$. This can be seen as the result of a transform (i.e., the *Bernstein transform* $\mathcal{B}_m$) mapping the sequence or vector of *Bernstein coefficients* $\boldsymbol{c} \in \mathbb{R}^{m+1}$ into the polynomial $p(x)$ in the variable $x \in [0, 1]$. Observe that the expected payoffs (25) and (27), and the private (30), external (31), and social (28) gain

[5] The link provided by Lemma 2 is however weaker, as conditions are necessary but not sufficient.

functions are all *polynomials in Bernstein form* in the mixed strategy x. The importance of this observation is that Bernstein transforms are endowed with many shape-preserving properties linking the sign patterns of the sequences of coefficients and the sign patterns of the respective polynomials (Farouki, 2012; Peña et al., 2014). We record some of the key properties that are relevant for our purposes in the following lemma. For more properties of Bernstein transforms see, e.g., Farouki (2012).

Lemma 3 (Properties of Bernstein transforms.)**.** Let $p(x) = \mathcal{B}_m(x; \boldsymbol{c})$ be a polynomial in Bernstein form of degree m with Bernstein coefficients $\boldsymbol{c}$. The Bernstein transform $\mathcal{B}_m$ satisfies:

1. **Lower and upper bounds.** For $x \in [0,1]$, the polynomial $p(x)$ satisfies the bounds $\min_{0 \le k \le m} c_k \le p(x) \le \max_{0 \le k \le m} c_k$.

2. **End-point values.** The initial and final points of $p(x)$ and $\boldsymbol{c}$ coincide, i.e., $p(0) = c_0$ and $p(1) = c_m$.

3. **Preservation of initial and final signs.** Let $\boldsymbol{c} \neq \mathbf{0}$. Then, the initial and final signs of $p(x)$ and $\boldsymbol{c}$ coincide, i.e., $\iota(p) = \iota(\boldsymbol{c})$ and $\phi(p) = \phi(\boldsymbol{c})$.

4. **Preservation of positivity.** The Bernstein transform of a positive (resp. negative) sequence is strictly positive (resp. strictly negative), i.e., if $\boldsymbol{c} \gneq \mathbf{0}$, then $p > 0$ (resp. if $\boldsymbol{c} \lneq \mathbf{0}$, then $p < 0$).

5. **Variation-diminishing property.** The number of sign changes of $p(x)$ is equal to the number of sign changes of $\boldsymbol{c}$ or less by an even amount, i.e., $\sigma(p) = \sigma(\boldsymbol{c}) - 2j$ where $j \geq 0$ is an integer.

6. **Derivatives.** The derivative of a polynomial in Bernstein form with coefficients $\boldsymbol{c}$ is proportional to a a polynomial in Bernstein form with coefficients $\boldsymbol{\Delta c}$. More precisely, we have

$$p'(x) = m \sum_{k=0}^{m-1} \binom{m-1}{k} x^k (1-x)^{m-1-k} \Delta c_k = m\mathcal{B}_{m-1}(x; \boldsymbol{\Delta c}), \tag{35}$$

   where $\Delta c_k = c_{k+1} - c_k$ is the first-forward difference of c_k.

7. **Preservation of sign patterns.** If the number of sign changes of $\boldsymbol{c}$ is at most one, then the sign patter of p coincides with the sign pattern of $\boldsymbol{c}$. That is: If $\sigma(\boldsymbol{c}) \leq 1$, then $\varrho(p) = \varrho(\boldsymbol{c})$.

Together with Eq. (29) (which links the social, private, and external gain functions) and Lemmas 1 and 2 (which link the sign pattern of gain functions to notions of individual and collective optimality), the properties of Bernstein transforms listed in Lemma 3 are the main tool we use to obtain our results.

3 What is a cooperative dilemma?

3.1 Two-player cooperative dilemmas with generic payoffs

In order to build our intuitions for the general multi-player case, we begin by considering the simple case of two players characterized in Example 1, with payoff matrix given by (8). In particular, we focus on the three prototypical types of two-player cooperative dilemmas we presented in Example 2, namely the prisoners' dilemma, the snowdrift (or chicken) game, and the stag hunt (or assurance game). We ask, as does Nowak (2012) for this same class of games: When can we say that action C corresponds to "cooperation" and action D to "defection"? And, relatedly: When can we say that the game represented by (8) is a "cooperative dilemma"? Although we ask similar questions, we arrive at different answers.

3.1.1 Preliminaries

Instead of adopting one existing definition, we start by looking at the commonalities among the three games, first at the level of their payoff orderings, and then at the level of their ESS structure in relation to the location of their social optima. We begin with the following observation.

Observation 1. The payoff orderings of the two-player prisoner's dilemma, the two-player snowdrift game, and the two-player stag hunt with generic payoffs are such that (ia) mutual C yields a higher payoff than mutual D, i.e., $P_1 > Q_0$ holds, (ib) players are always better off if their co-players play C than if they play D, i.e., $P_1 > P_0$ and $Q_1 > Q_0$ hold, and yet (ii) there is an individual incentive to play D, i.e., either $P_0 < Q_0$ or $P_1 < Q_1$ holds.

Conditions (ia) and (ib) can be regarded as the "benefits of cooperation", while condition (ii) can be regarded as the "costs of cooperation". The benefits of cooperation indicate not only (ia) that both players prefer mutual cooperation over mutual defection, but also (ib) that each player prefers their co-player to cooperate rather than to defect, or, in other words, that playing action C always induces a positive externality on the co-player. However, condition (ii) indicates that attempting to cooperate unilaterally can be costly, in the sense that it can lead to a less preferred individual outcome. In particular, if $P_0 < Q_0$ holds (as it happens in the prisoners' dilemma and the stag hunt, but not in the snowdrift game), defection yields a higher payoff than cooperation if the co-player defects, while if $P_1 < Q_1$ holds (as it happens in the prisoners' dilemma and the snowdrift game, but not in the stag hunt), defection yields a higher payoff than cooperation if the co-player cooperates. Both inequalities are satisfied for the prisoners' dilemma, making it the most stringent of the three cooperative dilemmas. In contrast, only one of the two inequalities of condition (ii) is satisfied for either the snowdrift game or the stag hunt, thus making these two games, in a sense, more "relaxed" cooperative dilemmas (Nowak, 2012).

What are the consequences of the payoff orderings of the prisoner's dilemma, the snowdrift game, and the stag hunt on their respective ESS structure and the location of their social optima? The sign patterns of the private gain sequence, the aggregate externality sequence, and the social gain sequence of these games have been characterized in Example 2. Moreover, their corresponding gain functions g (30), h (31), and f' (28) are, as for any other two-player game, linear in x. This allows to characterize their ESS structure and the location of their social optima in a straightforward way. In particular, the following result is easy to prove:

Lemma 4 (Two-player cooperative dilemmas with generic payoffs)**.** Consider the two-player cooperative dilemmas with generic payoffs introduced in Example 2.

1. The prisoners' dilemma has exactly one ESS, namely $x^* = 0$.
   (a) If $2P_1 \geq P_0 + Q_1$, the social optimum satisfies $\hat{x} = 1$.
   (b) If $2P_1 < P_0 + Q_1$, the social optimum $\hat{x}$ satisfies $0 < \hat{x} < 1$.
2. The snowdrift game has exactly one ESS $x^* \in (0, 1)$.
   (a) If $2P_1 \geq P_0 + Q_1$, the social optimum satisfies $\hat{x} = 1$.
   (b) If $2P_1 < P_0 + Q_1$, the social optimum $\hat{x}$ satisfies $0 < x^* < \hat{x} < 1$.
3. The stag hunt has two ESSs: $x_1^* = 0$ and $x_2^* = 1$. The social optimum satisfies $\hat{x} = 1$.

In the prisoners' dilemma, individual rationality or selection leads to an outcome where there is no cooperation ($x^* = 0$), although some cooperation is always socially optimal ($\hat{x} > 0$). In the snowdrift game, the only equilibrium features some cooperation ($0 < x^* < 1$) but it is always less than what is socially efficient ($x^* < \hat{x}$). In the stag hunt, the dilemma is one of coordination: while there is an efficient equilibrium with full cooperation that is socially optimal ($x_2^* = \hat{x} = 1$) there is also an inefficient one with nil cooperation ($x_1^* = 0$). Overall, each game features at least one inefficient equilibrium, and the level of cooperation sustained at such equilibrium is always lower than what would be socially optimal. We see this discrepancy between equilibria and social optima as the one capturing the tension between individual and collective interests that is the essence of any social and cooperative dilemma.

3.1.2 Definitions

We can now provide definite answers—in the specific context of two-player games with generic payoffs—to the questions of what is cooperation and what is a cooperative dilemma. We present these answers as definitions. We start by defining cooperation as an action that (i) benefits both players when they both play it, and (ii) benefits the co-player when a player plays it. More precisely, we have:

Definition 1 (Cooperation (two-player game with generic payoffs))**.** We say that action C of a two-player game with generic payoffs is cooperative if both (i) mutual C is preferred over mutual D, i.e., $P_1 > Q_0$ holds, and (ii) each player always prefers its co-player to play C than to play D, i.e., (iia) $P_1 > P_0$ and (iib) $Q_1 > Q_0$ hold.

Definition 1 essentially restates part (i) of Observation 1, that is, the "benefits of cooperation" part of our characterization of two-player cooperative dilemmas with generic payoffs. Our definition of cooperation agrees with the way Hauert et al. (2006, p. 196) define it implicitly for two-player games, but is otherwise in contrast with alternative definitions proposed in the literature. For instance, Nowak (2012) requires only condition (i), while Allen and Nowak (2015) seem to require, in their definition of "cooperative trait", condition (i) together with *either* (iia) or (iib). Allen and Nowak (2015) view conditions (iia) and (iib) as representing "different forms of help to the other player" and condition (i) as specifying that "this help is effective, in that it leads to a mutually beneficial outcome". Macy and Flache (2002) define cooperation implicitly (i.e., by stating conditions describing the "benefits of cooperation" of a cooperative dilemma) by requiring (i) and (iia) but replacing (iib) with the condition that "players prefer mutual cooperation over an equal probability of unilateral cooperation and defection", namely, that $2P_1 > P_1 + Q_0$ holds.

Out of the 24 different kinds of two-player games with generic payoffs (corresponding to the 24 possible strict orderings of the four payoff values), only five correspond to games with a cooperative action according to Definition 1. This is in contrast to the more generous definitions of cooperation by Nowak (2012) and Allen and Nowak (2015), according to which, respectively, twelve and eleven kinds of games correspond to games with a cooperative action. The five kinds of games picked by our definition include not only the prisoner's dilemma, the snowdrift game, and the stag hunt game, but also two additional ones, respectively characterized by rankings

$$P_1 > P_0 > Q_1 > Q_0, \tag{36}$$

and

$$P_1 > Q_1 > P_0 > Q_0. \tag{37}$$

It can be verified that for these two payoff orderings, the unique ESS $x^* = 1$ coincides with the social optimum $\hat{x} = 1$. We do not regard these games as capturing any dilemma, as individual and collective interests are perfectly aligned. To be more precise about this point, we introduce the following definition:

Definition 2 (Social dilemma)**.** A game is a social dilemma if it has an ESS x^* that is different from the social optimum $\hat{x}$.

Definition 2 is similar to the one given by Kollock (1998, p. 184), who defines a social dilemma as a game having "at least one deficient equilibrium". Yet, Kollock (1998) has in mind a (possibly asymmetric) pure-strategy NE as solution concept. In contrast, our analysis (i) is constrained to symmetric profiles, (ii) allows for mixed strategies, and (iii) is informed by evolutionary logic, and more specifically on the refinement of symmetric mixed NE given by the concept of ESS. Given the equivalence between ESS and ASE for two-strategy symmetric games, our solution concept picks those NE that are attractors of the replicator dynamic.

Building on Definitions 1 and 2 we are ready to define a cooperative dilemma as a game satisfying both the condition for having a cooperative action and the condition for being a social dilemma. That is, we have:

Definition 3 (Cooperative dilemma)**.** A game is a cooperative dilemma if (i) C is cooperative and (ii) the game is a social dilemma.

Part (i) of our definition captures the "benefits of cooperation" of a cooperative dilemma and is, again, for the two-player generic-payoff case, a restatement of part (i) of Observation 1. Part (ii) captures the "costs of cooperation", but they are stated in a different way: Cooperation is individually costly in the sense that individual incentives (or individual selection) can lead to an equilibrium that is inefficient, in the sense of being different from the social optimum. As it will be shown in Section 4, a necessary and sufficient condition for a game with a cooperative action to be a cooperative dilemma (and hence for the game to be a social dilemma) is that, in addition, there are incentives to defect in a specific sense, generalizing part (ii) of Observation 1.

Our definition picks the prisoner's dilemma, the snowdrift game, and the stag hunt game as the only cooperative dilemmas among the 24 different two-player games with generic payoffs. In contrast, previous definitions of two-player cooperative dilemmas (Hauert et al., 2006; Nowak, 2012; Allen and Nowak, 2015) are more generous. For instance, and in the context of two-player games, Hauert et al. (2006) defines "social dilemmas" as games satisfying all the conditions of Definition 1 together with

$$Q_1 > P_0, \tag{38}$$

i.e., the requirement that "in any mixed group defectors outperform cooperators", which they interpret as the "costs of cooperation". As a result, they classify as cooperative dilemmas four different games: the three cooperative dilemmas that we identify plus a game of "by-product mutualism", which corresponds to the payoff ranking (37). As we have explained, the only ESS of such game is $x^* = 1$, which coincides with the social optimum $\hat{x} = 1$ and is not a social dilemma according to Definition 2. Hauert et al. (2006) are well aware of this, as they comment that in this case "[t]he dilemma is completely relaxed". Nowak (2012) defines a cooperative dilemma as a game satisfying, first, condition (i) of Definition 1 (the "benefits of cooperation") and second, either part (ii) of Observation 1 or condition (38) (the "costs of cooperation"). This results in a very broad definition of a "cooperative dilemma", which includes eight of the 24 different two-player games. Allen and Nowak (2015) revisit this definition of cooperative dilemma by enlarging the "benefits of cooperation" to also include either part (iia) or (iib) of Definition 1. The games classified as "social dilemmas" according to this seemingly more restrictive definition are however the same as those following Nowak (2012)'s original definition.

3.2 Multi-player cooperative dilemmas

Having picked satisfactory definitions of cooperation and cooperative dilemmas for the simple case of two-player games with generic payoffs, we take a broader perspective and ask: How should we generalize the definitions given in Section 3.1 to encompass also multi-player games with possibly non-generic payoffs? Since the definition of a social dilemma given in Definition 2 is already general, our problem is more precisely how to expand Definition 1 of a cooperative action so that it covers also the more general case. Once this generalization is obtained, we can continue using Definition 3 of a cooperative dilemma as a game with a cooperative action that is a social dilemma.

We start with the straightforward part. Moving from $n = 2$ players to $n \geq 2$ players, we generalize part (i) of Definition 1 as follows.

Definition 4. We say that universal C is preferred over universal D if

$$P_{n-1} > Q_0 \tag{39}$$

holds.

Condition (39) simply means that players obtain a larger payoff if they all choose C than if they all choose D. This condition is often encountered as part of the "benefits of cooperation" of previous definitions of multi-player "social dilemmas", "cooperative dilemmas" or "cooperation games", and hence implicitly included as a property of a cooperative action (Dawes, 1980; Nowak, 2012; Rand and Nowak, 2013; Hilbe et al., 2014; Peña et al., 2016; Płatkowski, 2017).

The generalization of part (ii) of Definition 1 is less straightforward. For $n = 2$, the additional requirement for C to be cooperative is that the switch from D to C by a focal player results in a positive externality on the co-player. For $n > 2$, there is more than one co-player and thus different ways of understanding what a positive externality on co-players might mean. Recalling our switching experiment for $n \geq 2$ (see Section 2.2), we can think of two alternatives. First, it might be required that the switch of the focal player makes the other players, taken as a block, never worse off (and at least sometimes better off). Second, a stronger condition might be required, namely that the switch makes each of the co-players, taken individually, never worse off (and at least sometimes better off). To be more precise, we have the following definitions in terms of the sequences we have introduced in Section 2.[6]

Definition 5 (Positive aggregate externalities)**.** We say that action C induces positive aggregate externalities if the aggregate externality sequence is positive, i.e., if

$$\boldsymbol{E} \gneq \boldsymbol{0} \tag{40}$$

holds.

[6] It is immediate from Eq. (5) that requiring positive individual externalities in the sense of Definition (6) is indeed stronger than requiring positive aggregate externalities in the sense of Definition (5), i.e., positive individual externalities imply positive aggregate externalities.

Definition 6 (Positive individual externalities)**.** We say that action C induces positive individual externalities if both payoff sequences are non-decreasing and at least one is increasing, i.e., if both

$$\Delta \boldsymbol{P} \geq \mathbf{0} \text{ and } \Delta \boldsymbol{Q} \geq \mathbf{0}, \tag{41a}$$

$$\Delta \boldsymbol{P} \gneq \mathbf{0} \text{ or } \Delta \boldsymbol{Q} \gneq \mathbf{0} \tag{41b}$$

hold.

Both stronger and weaker versions of conditions (40) and (41) (and the underlying concepts of positive aggregate and individual externalities) have previously appeared in the literature to characterize the "benefits of cooperation" (and hence the meaning of a cooperative action) in population genetics and game-theoretic models. First, a stronger version of (40) (namely that the aggregate externality sequence is strictly positive, $\boldsymbol{E} > \mathbf{0}$) appears as part of the "focal-complement" interpretation of altruism proposed by Kerr et al. (2004) (based on previous work by Matessi and Karlin, 1984). Second, a stronger version of (41) (namely that the payoff sequences are both strictly increasing, $\Delta \boldsymbol{P} > \mathbf{0}$ and $\Delta \boldsymbol{Q} > \mathbf{0}$) appears as part of the "individual-centered" interpretation of altruism proposed by Kerr et al. (2004) (based on previous work by Uyenoyama and Feldman, 1980). Third, and finally, a weaker version of condition (41) (namely that the payoff sequences are both non-decreasing, i.e., Eq. (41a) without the additional requirement in Eq. (41b)) appears as part of the definitions of "n-player social dilemmas" (Hilbe et al., 2014), "cooperation games" (Peña et al., 2016), and "multi-player social dilemmas" (Płatkowski, 2017).

For $n = 2$, the conditions for positive aggregate externalities and positive individual externalities given in Definitions 5 and 6 are equivalent, as in this case there is only one co-player per player. Indeed, in the two-player case the aggregate externality sequence reduces to $\boldsymbol{E} = (Q_1 - Q_0, P_1 - P_0)$ (see Example 1), so that conditions (40) and (41) both simplify to $P_1 \geq P_0$ and $Q_1 \geq Q_0$, with at least one inequality being strict. However, Definitions 5 and 6 are different for $n > 2$, with positive individual externalities implying positive aggregate externalities, but not vice versa, as it can be verified from Eq. (5). With the aim of being as general as possible, we choose the condition of positive aggregate externalities over the condition of positive individual externalities to be part of our definition of cooperative action. Thus, we arrive at:

Definition 7 (Cooperation)**.** We say that action C of a multi-player game is cooperative if both (i) universal C is preferred over universal D, and (ii) C induces positive aggregate externalities.

Definition 7 generalizes the conditions for cooperation given by Definition 1 to multi-player games with possibly non-generic payoffs in a relatively inclusive way. Similarly, when taken as part of Definition 3, it expands the definition of cooperative dilemmas to include also interactions among multiple players and the possibility of non-generic payoffs.

Below, we illustrate Definition 7 by showing how the public goods games of Example 3, and the participation games of Examples 8 and 9 all fall into the category of games with a cooperative action. We postpone showing how these examples also fall into the category of cooperative dilemmas until the next section.

Example 3 (continued)**.** Since $\boldsymbol{E}$ is positive, C (contributing) induces positive aggregate externalities. Moreover, since both $\boldsymbol{P}$ and $\boldsymbol{Q}$ are increasing, C also induces positive individual externalities. Since, additionally, $b_{n-1} - b_0 > c_n$ holds, then $P_{n-1} > Q_0$ holds, and action C is cooperative.

Example 8 (continued)**.** Since $\Delta \boldsymbol{P} \geq \boldsymbol{0}$ and $\Delta \boldsymbol{Q} \gneq \boldsymbol{0}$ hold, C induces both positive individual and aggregate externalities. Additionally, since $P_{n-1} = \gamma > v_{n-1} = Q_0$ also holds, action C (staying "out") is cooperative.

Example 9 (continued)**.** Since $\Delta \boldsymbol{P} \gneq \boldsymbol{0}$ and $\Delta \boldsymbol{Q} \geq \boldsymbol{0}$ hold, C induces both positive individual and aggregate externalities. Additionally, since $P_{n-1} = v_n > \gamma = Q_0$ also holds, action C (choosing "in") is cooperative.

Examples 3, 8, and 9 are such that action C induces both positive individual externalities and positive aggregate externalities. The following two examples illustrate games for which the cooperative action does not necessarily induce positive individual externalities.[7]

Example 10 (Competition with a superior choice)**.** Consider the congestion game put forward by Menezes and Pitchford (2006). Individuals choose between two alternative choices C and D, such as physical locations, product spaces, roads, or bars. There is competition (or congestion) as individual payoffs fall when more players make the same choice. It follows that $\boldsymbol{P}$ is decreasing and $\boldsymbol{Q}$ is increasing. Since $\boldsymbol{P}$ is decreasing, action C does not induce positive individual externalities. Let us assume, as do Menezes and Pitchford (2006), that C is "superior", in the sense that all players prefer C to D if the same number of players choose C or D, e.g., bar C offers better music (or simply has more tables) than bar D. This implies that $P_k > Q_{n-1-k}$ holds for all $k = 0, 1, \ldots, n-1$, and, in particular, that $P_{n-1} > Q_0$ holds. Hence, universal C is preferred over universal D. Additionally, note that $\boldsymbol{E} \gneq \boldsymbol{0}$ can hold, provided that the switch from D to C by a focal player is such that the positive externality due to decreased competition experienced by all other D-players compensates for the negative externality due to increased competition experienced by all other C-players. In this case, C is cooperative according to our definition but, as stated above, does not induce positive individual externalities.

Example 11 (Majority game with superior choice)**.** Consider a "majority game" among an odd number of players (i.e., $n = 2m + 1$ with m an integer greater than zero). There are two choices (e.g., policies, candidates) that individuals can vote over: C and D. The option with more votes gets selected (majority rule). Voting is costless. All players obtain a payoff of zero

[7] For yet another example related to public goods provision, see the model of "antisocial rewarding" analyzed by dos Santos and Peña (2017, p. 8).

if the option they have chosen is not selected. C-players (resp. D-players) obtain a payoff of $\alpha > 0$ (resp. $\beta > 0$) if their option is selected. The payoffs are then given by

$$P_k = \alpha [\![k \geq m]\!], \ k = 0, 1, \dots, n-1 \tag{42a}$$

$$Q_k = \beta [\![k \leq m]\!], \ k = 0, 1, \dots, n-1. \tag{42b}$$

With this specification, $\boldsymbol{P}$ is increasing but $\boldsymbol{Q}$ is decreasing. Hence C does not induce positive individual externalities (nor does D). However, C induces positive aggregate externalities whenever $\alpha > \beta$ holds (i.e., when C-players' preference for their choice is larger than D-players' preference for their choice). Indeed, the external gains are given by

$$E_k = (\alpha - \beta) [\![k = m]\!], \ k = 0, 1, \dots, n-1, \tag{43}$$

and hence $\boldsymbol{E} \gneqq \boldsymbol{0}$ holds. Since, additionally, $P_{n-1} = \alpha > \beta = Q_0$ holds, universal C is preferred over universal D, and action C is cooperative.

4 When is a game a cooperative dilemma?

We have defined a cooperative dilemma (Definition 3) as a game with a cooperative action (Definition 7) that is also a social dilemma (Definition 2). In this section, we look into conditions for a game to be a cooperative dilemma (and hence a social dilemma) that can be verified without the need for checking directly whether or not there exists at least one ESS x^* such that $x^* \neq \hat{x}$. We first provide a general necessary and sufficient condition in terms of the private gain function. Then we provide simpler conditions given solely in terms of the private gain sequence. These are necessary and sufficient for two players but not in general.

4.1 Necessary and sufficient condition

We begin by noting and recording two simple consequences of action C being cooperative. First, since universal C is preferred over universal D if action C is cooperative, it must follow that the social optimum is greater than zero, i.e., that some cooperation is required to maximize the expected population payoff. We record this simple observation in the following lemma.

Lemma 5. Suppose universal C is preferred over universal D. Then $\hat{x} > 0$ holds.

Proof. If universal C is preferred over universal D then, by Definition 4 and the end-point values property of Bernstein transforms (see Lemma 3.2), $f(1) = f_C(1) = P_{n-1} > Q_0 = f_D(0) = f(0)$ holds, implying that $x = 0$ does not maximize the expected payoff f. Hence $\hat{x} \neq 0$ and thus $\hat{x} > 0$ holds. ☐

Second, if C is cooperative, then it induces positive aggregate externalities, i.e., the aggregate externality sequence must be positive. This implies (by the preservation of positivity property of Bernstein transforms, see Lemma 3.5) that the external gain function is positive. Thus, we have:

Lemma 6. Suppose that C induces positive aggregate externalities. Then $h > 0$ holds.

Lemma 6 implies (via identity (29), and hence $g(x) = f'(x) - h(x)$) that the private gain function is strictly smaller than the social gain function. Using Lemmas 5 and 6 together with Lemma 1, allows us to prove the following result.

Proposition 1. Suppose C is cooperative. Then the game is a cooperative dilemma if and only if there exists $x \in [0, 1]$ such that $g(x) < 0$, or, equivalently, if and only if $x^* = 1$ is not its only ESS.

Proof. Let C be cooperative. Then, by Lemmas 5 and 6, both $\hat{x} > 0$ and $h > 0$ hold. Using these observations, we can prove the proposition by considering the following three exhaustive cases.[8]

1. If g is negative (i.e., $g \lneq 0$), then there exist $x \in [0, 1]$ such that $g(x) < 0$. Further, by Lemma 1, $x^* = 0$ is its unique ESS. As $\hat{x} > 0$ holds, the game is thus a social dilemma and, therefore, a cooperative dilemma.

2. If g changes sign at least once (i.e., $\sigma(g) \geq 1$), there exists x such that $g(x) < 0$. We have the following two cases.

   (a) If the initial sign of g is negative (i.e., $\iota(g) = -1$), then, by Lemma 1, $x^* = 0$ is an ESS. As $\hat{x} > 0$ holds, the game is a social dilemma and, therefore, a cooperative dilemma.

   (b) If the initial sign of g is positive (i.e., $\iota(g) = 1$), then, by Lemma 1, there exists at least one interior ESS $x^* \in (0, 1)$, which then satisfies the condition $g(x^*) = 0$. As $h > 0$ holds, we have $h(x^*) > 0$, which implies $f'(x^*) > 0$ via identity (29). As x^* is interior, this implies $x^* \neq \hat{x}$. Hence, the game is a social dilemma and, therefore, a cooperative dilemma.

3. If g is positive (i.e., $g \gneq 0$), then there does not exist $x \in [0, 1]$ such that $g(x) < 0$. Further, by Lemma 1, $x^* = 1$ is the unique ESS. In addition, since $h > 0$ holds, $f' > 0$ holds. Hence, $\hat{x} = 1$. Since the unique ESS coincides with the social optimum, the game is not a social dilemma and, therefore, not a cooperative dilemma.

□

Proposition 1 provides a necessary and sufficient condition for characterizing a cooperative dilemma, namely, that the private gain function is negative for at least some value of its domain. In other words, players must have an ex ante incentive (in terms of their private gains in expected payoff) to choose D over C for at least some symmetric mixed-strategy profile played by co-players.

[8] Our assumption $\boldsymbol{P} \neq \boldsymbol{Q}$, which implies $\boldsymbol{G} \neq 0$, precludes the case where $g(x) = 0$ holds for all $x \in [0, 1]$.

4.2 Simpler conditions

It is convenient to have simpler conditions to check if a game is a cooperative dilemma or not. The properties of Bernstein transforms provide us with such conditions, which we state in the following corollaries to Proposition 1.

Corollary 1. Suppose C is cooperative. If either $\iota(\boldsymbol{G}) = -1$ or $\phi(\boldsymbol{G}) = -1$, then the game is a cooperative dilemma.

Proof. By the preservation of initial and final signs (Lemma 3.3) $\iota(\boldsymbol{G}) = -1$ implies $\iota(g) = -1$ and $\phi(\boldsymbol{G}) = -1$ implies $\phi(g) = -1$. In either case the existence of x such that $g(x) < 0$ is implied, so that the result follows from Proposition 1. □

Corollary 2. Suppose C is cooperative. If the game is a cooperative dilemma, then $G_k < 0$ holds for some $k = 0, 1 \ldots, n-1$.

Proof. If $G_k \geq 0$ holds for all $k = 0, 1 \ldots, n-1$, then preservation of positivity (Lemma 3.4) implies $g \geq 0$. Hence, Proposition 1 implies that $G_k < 0$ holds for some $k = 0, 1 \ldots, n-1$ if the game is a cooperative dilemma. □

Corollaries 1 and 2 offer straightforward criteria to determine whether a game in which C is a cooperative action qualifies as a dilemma or not. While Corollary 1 gives a sufficient condition, Corollary 2 gives a necessary condition. The condition in Corollary 1 is that either the initial or the final sign of the private gain sequence is negative. In other words, if there is an individual incentive to defect when either sufficiently many co-players are cooperating or sufficiently many co-players are defecting, then the game is a cooperative dilemma. The condition in Corollary 2 is that, for the game to be a cooperative dilemma, there must be some ex post incentives to defect. For two players ($n = 2$), the conditions in the two corollaries are equivalent and simplify, if payoffs are generic, to condition (ii) of Observation 1. For more players ($n > 2$) the conditions do not coincide and their converses are not true. Thus, there are cooperative dilemmas where both the initial and the final signs of $\boldsymbol{G}$ are positive as well as games where $G_k < 0$ holds for some $k = 0, 1, \ldots, n-1$ but that are not cooperative dilemmas.

We illustrate Corollary 1 by showing how it implies that (under appropriate additional assumptions) our previously considered Examples 3, 8, and 9 are cooperative dilemmas.

Example 3 (continued)**.** We have shown that contributing to the public good (playing C) is cooperative. We have also shown that, as long as costs are sufficiently high, the initial or the final sign of the public goods games discussed in Examples 4, 5, 6, and 7 are negative. Hence Corollary 1 applies, and these games are cooperative dilemmas.

Example 8 (continued)**.** Since not participating (playing C) is cooperative and $\varrho(\boldsymbol{G}) = (1, -1)$ holds, congestion games are cooperative dilemmas by Corollary 1.

Example 9 (continued)**.** Since participating (playing C) is cooperative and $\varrho(\boldsymbol{G}) = (-1, 1)$ holds, games with participation synergies are cooperative dilemmas by Corollary 1.

We defer until Section 6.1 showing that Examples 10 and 11 are also cooperative dilemmas.

5 When is universal cooperation socially optimal?

We have seen that the social optimum satisfies $\hat{x} > 0$ for all cooperative dilemmas, i.e., some level of cooperation is required to maximize the expected payoff (see Lemma 5). It is often of interest to distinguish between cooperative dilemmas satisfying $\hat{x} = 1$ and those satisfying only $0 < \hat{x} < 1$. Indeed, some authors (e.g., Macy and Flache (2002)) would argue that only the first group satisfies the conditions of a cooperative dilemma. This motivates the following definition.

Definition 8 (Social optimality of universal C)**.** We say that universal C is socially optimal if $\hat{x} = 1$ holds.

For two-player games it is straightforward to determine whether universal C is socially optimal or not. In particular, for the prototypical two-player cooperative dilemmas with generic payoffs considered in Example 2 the condition $2P_1 \geq P_0 + Q_1$ is both necessary and sufficient for the social optimality of universal C. For the cases of the prisoners' dilemma and the snowdrift game this is immediately apparent from the statement of Lemma 4; for the case of the stag-hunt the claim follows from Lemma 4.3 upon noting that the payoff inequalities defining a stag hunt in Example 2.3 imply $2P_1 > P_0 + Q_1$.

For an arbitrary number of players n, Lemma 2.2 has identified a positive final sign of the social gain function f' as a necessary condition for the social optimality of universal C. While we offer a (slight but useful) refinement of this condition in Proposition 3 below, our main interest in this section is in providing simple, general sufficient conditions for the social optimality of universal C. For our purposes having such conditions is of particular interest because once we know that universal C is socially optimal in a cooperative dilemma, we can immediately conclude that cooperation is underprovided at any inefficient ESS of such a game. In contrast, if universal C is not socially optimal in a cooperative dilemma, so that $0 < \hat{x} < 1$ holds, the possibility that an inefficient ESS x^* of such a game features overprovision of cooperation, i.e. $x^* > \hat{x}$ holds, can no longer be dismissed a priori—an issue to which we return in Section 6.

5.1 Positive social gains and the social optimality of universal C.

We begin with a definition.

Definition 9 (Positive social gains)**.** We say that action C induces positive social gains if the social gain sequence is positive, i.e., if

$$\boldsymbol{S} \gneq \boldsymbol{0}. \tag{44}$$

According to Definition 9, action C induces positive social gains if, as a result of our switching experiment, the switch of a focal player from playing D to playing C increases the total payoff of all players (including the focal). Definition 9 is thus related to Definitions 5 and 6 of positive aggregate and individual externalities. As discussed in Section 3.2, conditions related to those appearing in the statements of these definitions have been previously used as characterizing the "benefits of cooperation" in cooperative dilemmas by different authors, and in particular as part of different interpretations of altruism in the population genetics literature reviewed by Kerr et al. (2004). Condition (44) is no exception: a stronger version of it, namely that the social gain sequence is strictly positive, i.e., $\boldsymbol{S} > \mathbf{0}$, appears as part of the "multilevel" interpretation of altruism proposed by Kerr et al. (2004), based on previous work by Matessi and Jayakar (1976) and Cohen and Eshel (1976), among others.

It is intuitive that universal C should be socially optimal if it is the case that, no matter which pure-strategy profile we consider, switching the action of a single player from D to C never decreases but sometimes increases the total payoff, that is, if the action C induces positive social gains. The proof of the following proposition verifies this intuition; thereafter we illustrate its application to two of our previous examples.

Proposition 2. Suppose that action C induces positive social gains. Then universal C is socially optimal.

Proof. From the preservation of positivity of Bernstein transforms (Lemma 3.4), $\boldsymbol{S} \gneq \mathbf{0}$ implies $f' > 0$. Consequently, the expected payoff $f(x)$ is strictly increasing in x, implying $\hat{x} = 1$. □

Example 3 (continued)**.** Suppose that condition (14) holds. This is the case, for instance, if there is "cost sharing" (e.g., Weesie and Franzen (1998)), i.e., if the cost sequence is given by $c_i = \gamma/i$ for some constant $\gamma > 0$. Then, as we have discussed before, $\boldsymbol{S} \gneq \mathbf{0}$ holds, i.e., C induces positive social gains. It follows by Proposition 2 that universal C is socially optimal.

As a partial counterpart to the sufficient condition for the social optimality of universal C in Proposition 2 we have the result that for universal C to be socially optimal it must be that the final sign of $\boldsymbol{S}$ is positive.

Proposition 3. Suppose that universal C is socially optimal. Then $\phi(\boldsymbol{S}) = 1$ holds.

Proof. From Lemma 2.2, a necessary condition for $\hat{x} = 1$ is that the final sign of f' is positive, i.e., that $\phi(f') = 1$ holds. Since f' is the Bernstein transform of the social gain sequence $\boldsymbol{S}$, and by the preservation of initial and final signs of Bernstein transforms (Lemma 3.3), this is equivalent to requiring that $\phi(\boldsymbol{S}) = 1$ holds. □

Of course, Proposition 3 implies that whenever the final sign of the social gain sequence is negative, the social optimum satisfies $\hat{x} < 1$.

Example 7 (continued)**.** Suppose that $1 < \theta < n$, i.e., the threshold public goods game is a teamwork dilemma. Then, by substituting (16) and $c_i = \gamma$ into (13), we obtain $S_{n-1} = -\gamma < 0$. The final sign of the social gain sequence is thus negative and the social optimum satisfies $\hat{x} < 1$.

It is noteworthy that for two-player cooperative dilemmas and generic payoffs, the condition $\phi(\boldsymbol{S}) = 1$ is not only necessary, as established in Proposition 3, but also sufficient for the social optimality of universal C. To see this, observe that for every two-player game we have (Example 1) $S_{n-1} = 2P_1 - P_0 - Q_1$ so that the condition $S_{n-1} \geq 0$, which is in turn implied by $\phi(\boldsymbol{S}) = 1$, is sufficient for the social optimality of C in any one of the prototypical two-person cooperative dilemmas. For $n > 2$ a similar conclusion is not possible: there are cooperative dilemmas where $x = 1$ is a local maximizer of total fitness $f(x)$, so that $\phi(\boldsymbol{S}) = 1$ holds, but universal C is not optimal because $x = 1$ is not a global maximizer. See the following example for an illustration.

Example 12. Consider the three-player game with payoff sequences given by $\boldsymbol{P} = (0, 1, 1 + z)$ and $\boldsymbol{Q} = (1, 2, 1)$, with $0 < z < 1/3$. The private gains are then given by $\boldsymbol{G} = (-1, -1, z)$, the aggregate externalities by $\boldsymbol{E} = (2, 0, 2z)$, the social gains by $\boldsymbol{S} = (1/3, -1/3, z)$, and the total payoffs by $\boldsymbol{T} = (3, 4, 3, 3(1 + z))$. Since $P_2 = 1 + z > 1 = Q_0$ and $\boldsymbol{E} \gneq \boldsymbol{0}$, the game is such that C is a cooperative action. Further, the initial sign of the private gain sequence is negative, i.e., $\iota(\boldsymbol{G}) = -1$. Hence, the game is a cooperative dilemma by Corollary 1. Moreover, $\phi(\boldsymbol{S}) = 1$ holds, so $x = 1$ locally maximizes the expected payoff $f(x)$. However, $x = 1$ is not a global maximizer if z is sufficiently small. This is illustrated in Fig. 1 for $z = 1/10$. In this case, $\hat{x} = 0.352527$.

5.2 Alternative conditions for the social optimality of universal C

The condition that C induces positive social gains in Proposition 2 is equivalent to requiring that the total payoff sequence $\boldsymbol{T}$ is increasing, thereby ensuring that n maximizes the total payoff T_i over the number i of players choosing C. The following lemma shows that this weaker requirement does suffice to ensure the social optimality of universal C.

Lemma 7. Suppose that $T_n = \max_{0 \leq i \leq n} T_i$. Then universal C is socially optimal.

Proof. The expected payoff f is the Bernstein transform of the average payoff sequence $\boldsymbol{T}/n = (T_0/n, \ldots, T_n/n)$ (see Eq. (27)). By the lower and upper bounds and end-point values properties of polynomial in Bernstein form (see numerals 1 and 2 in Lemma 3), it then follows that

$$f(1) = \frac{T_n}{n} = \max_{0 \leq i \leq n} \frac{T_i}{n} \geq f(x) \tag{45}$$

holds for all $x \in [0, 1]$. Hence $x = 1$ maximizes $f(x)$. Using our assumption that the social optimum $\hat{x}$ is unique, $\hat{x} = 1$ follows. □

Lemma 7 makes intuitive sense: If the sum of payoffs to players playing pure strategies is maximized when all players choose C, it follows that the pure strategy $x = 1$ maximizes expected payoff and is hence the social optimum. One might think that the condition $T_n = \max_{0 \leq i \leq n} T_i$ is also necessary for the social optimality of universal C. However, for $n > 2$ this is not so: $\hat{x} = 1$ does not imply that T_n maximizes the total payoffs (i.e., the converse of Lemma 7 is not

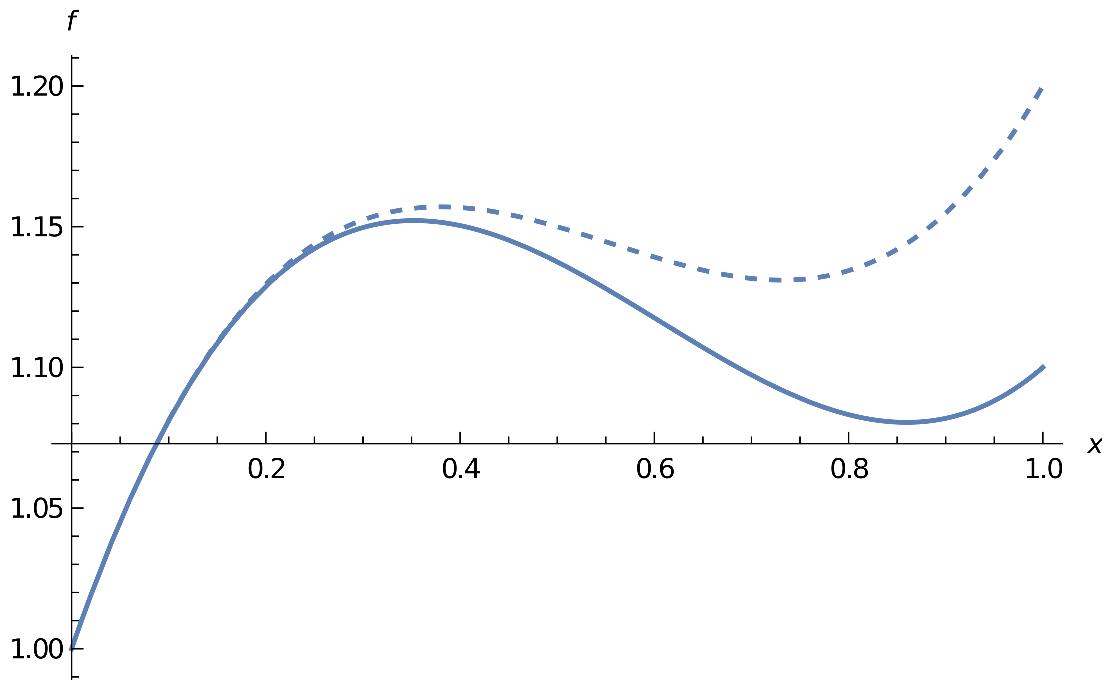

Figure 1: Expected payoff $f(x)$ for the game in Example 12 for two values of z. For $z = 1/5$ (*dashed line*), the social optimum satisfies $\hat{x} = 1$ and coincides with the ESS at $x_2^* = 1$. For $z = 1/10$ (*solid line*), the social optimum satisfies $\hat{x} = 0.352527$. In this case the social optimum is below the ESS $x_2^* = 1$. Such an ESS then features "too much cooperation".

true). In particular, determining whether or not universal C is socially optimal in cooperative dilemmas where both $\phi(\boldsymbol{S}) = 1$ and $\max_{0 \le i \le n} T_i \neq T_n$ hold is a non-trivial task in general, and requires additional assumptions on the structure of the cooperative dilemma.

Example 12 (continued)**.** In this example both $\phi(\boldsymbol{S}) = 1$ and T_1 maximizes the total payoff sequence for all values of $z \in (0, 1/3)$, so that $\max_{0 \le i \le 3} T_i = T_1 \neq T_3$ holds. Whether or not universal C is socially optimal depends on the value of the parameter z. For sufficiently low z, the social optimum satisfies $\hat{x} < 1$ (and universal C is not socially optimal) while for sufficiently high z, the social optimum satisfies $\hat{x} = 1$ (and universal C is not socially optimal). See Fig. 1 for an illustration with $z = 1/10$ (low z) and $z = 1/5$ (high z).

We conclude this section by applying Lemma 7 to obtain a very simple sufficient condition for the social optimality of C for games in which C is not only cooperative but also induces positive individual externalities.

Proposition 4. Suppose C is cooperative and induces positive individual externalities. If $G_{n-1} \ge 0$ holds, then universal C is socially optimal.

Proof. As C is cooperative, we have $T_n = nP_{n-1} > nQ_0 = T_0$. By Lemma 7 it then suffices to show that $T_n \ge T_i$ holds for all $i = 1, \ldots, n-1$ to prove the result.

The condition $T_n \ge T_i$ is equivalent to

$$T_n - T_i = nP_{n-1} - iP_{i-1} - (n-i)Q_i \ge 0.$$

Because $\boldsymbol{P}$ is non-decreasing by the assumption of positive individual externalities, we have

$$nP_{n-1} - iP_{i-1} \ge (n-i)P_{n-1},$$

so that the desired inequality follows if $P_{n-1} - Q_i \ge 0$ holds. Because $\boldsymbol{Q}$ is non-decreasing, this in turn is implied by the assumption $G_{n-1} \ge 0$, which is equivalent to $P_{n-1} - Q_{n-1} \ge 0$. $\square$

We have seen that the public goods games with convex benefits and intermediate costs (Example 5), the public goods games with sigmoid benefits and intermediate costs such that $\Delta b_0 < \gamma \le \Delta b_{n-1}$ (Example 6, which includes the case of Example 7 with $\theta = n$), and the games with participation synergies (Example 9) are all such that C is cooperative and that C induces positive individual externalities. Morover, the final sign of the private gain sequences of these games is positive, which implies $G_{n-1} \ge 0$. By an application of Proposition 4, we can then conclude that in these games universal C is socially optimal, i.e., $\hat{x} = 1$ holds.

6 Multi-player prisoners' dilemmas, snowdrift games, and stag hunts

6.1 Definitions

Consider again the two-player cooperative dilemmas we discussed in Section 3.1, and their possible generalization to more than two players. When is it appropriate to call an n-player game a prisoners' dilemma, a snowdrift game, or a stag hunt? As a first step to answer this question, we would like definitions of these terms that (i) satisfy the conditions of a cooperative dilemma as defined in Definition 3, (ii) include the corresponding two-player generic-payoff versions as particular cases, and (iii) are minimal and given solely in terms of inequalities involving simple operations on the payoff sequences $\boldsymbol{P}$ and $\boldsymbol{Q}$. These considerations lead us to the following definition.

Definition 10 (Prisoners' dilemmas, snowdrift games, and stag hunts)**.** Let C be cooperative, and let $\varrho(\boldsymbol{G})$ denote the sign pattern of the private gain sequence $\boldsymbol{G}$.

1. We say that the game is a prisoners' dilemma if $\varrho(\boldsymbol{G}) = (-1)$, i.e., if $\boldsymbol{G} \lneq \boldsymbol{0}$.

2. We say that the game is a snowdrift game if $\varrho(\boldsymbol{G}) = (1, -1)$ i.e., $\boldsymbol{G}$ has a single sign change from positive to negative.

3. We say that the game is a stag hunt if $\varrho(\boldsymbol{G}) = (-1, 1)$, i.e., $\boldsymbol{G}$ has a single sign change from negative to positive.

We have so far seen several examples of these three classes of multi-player cooperative dilemmas. First, when contribution costs are sufficiently high, all public goods games of Examples 4, 5, and 6 have a private gain sequence $\boldsymbol{G}$ that is negative and thus qualify as prisoners' dilemmas. Second, the public goods game with concave benefits and fixed intermediate costs of Example 4, the public goods game with sigmoid benefits and fixed intermediate costs such that $\Delta b_{n-1} < \gamma \leq \Delta b_0$ of Example 6 (which includes the volunteer's dilemma defined in Example 7), and the congestion games of Example 8 have all a private gain sequence with pattern $\varrho(\boldsymbol{G}) = (1, -1)$ and are thus, according to our definition, particular instances of snowdrift games. Third, and lastly, the public goods games with convex benefits and intermediate costs of Example 5, and the public goods games with sigmoid benefits and intermediate costs such that $\Delta b_0 < \gamma \leq \Delta b_{n-1}$ of Example 6 (which include the case of Example 6 with $\theta = n$), and the games with participation synergies of Example 9 have all a private gain sequence characterized by sign pattern $(-1, 1)$ and are thus particular instances of stag hunts. Regarding Examples 10 and 11 we have the following characterization.

Example 10 (continued)**.** Since $\boldsymbol{P}$ is decreasing and $\boldsymbol{Q}$ is increasing by the assumption of competition, it is clear that $\Delta\boldsymbol{G} = \Delta\boldsymbol{P} - \Delta\boldsymbol{Q} \lneq \boldsymbol{0}$ holds, so that $\boldsymbol{G}$ is decreasing. Additionally, $P_0 > Q_0$ (i.e., being alone at C rather than being with $n-1$ other players at D) follows from the assumptions that C is superior together with the assumption of competition, as $P_0 > Q_{n-1} \geq Q_{n-2} \geq \ldots \geq Q_0$ holds (Menezes and Pitchford, 2006). Now assume, as do Menezes and Pitchford (2006), that being alone at D (which gives a payoff of Q_{n-1}) is better than being at C and competing with everyone else (which gives a payoff of P_{n-1}). Then, $P_{n-1} < Q_{n-1}$ holds. It follows that the private gain sequence has a single sign change from positive to negative, i.e., $\varrho(\boldsymbol{G}) = (1, -1)$ holds. This, together to the fact (that we have shown before) that C is cooperative, allows us to conclude that the game is an instance of a snowdrift game, as specified in Definition 10.2.

Example 11 (continued)**.** By substituting (42) into the definition of private gains (4) we obtain

$$G_k = \alpha[\![k > m]\!] + (\alpha - \beta)[\![k = m]\!] - \beta[\![k < m]\!],\ k = 0, 1, \ldots, n-1. \tag{46}$$

Since $\alpha > 0$ and $\beta > 0$, the private gain sequence has a single sign change from negative to positive, i.e., $\varrho(\boldsymbol{G}) = (-1, 1)$ holds. Moreover, C is also cooperative. Thus, according to Definition 10.3, the game is a stag hunt.

For all of the games in Definition 10, either the initial or the final sign of the private gain sequence is negative. It then follows from Corollary 1 that all games in Definition 10 constitute cooperative dilemmas. Further, for $n = 2$ the private gain sequence cannot have more than one sign change, so that the only case of a game with cooperative action C not covered by Definition 10 is the one in which $\boldsymbol{G} \gneq \boldsymbol{0}$ holds. From Corollary 2 we thus obtain:

Corollary 3. Let $n = 2$. Then a game is a cooperative dilemma if and only if it is a prisoner's dilemma, snowdrift game or stag hunt in the sense of Definition 10.

For the case of two players ($n = 2$), Definition 10 expands the scope of two-player prisoners' dilemmas, snowdrift games, and stag hunt games as previously defined in Section 3.1 to include also non-generic cases. For example, the games characterized by payoff orderings $Q_1 = P_1 > Q_0 > P_0$ and $Q_1 > P_1 > Q_0 = P_0$ are also classified as prisoners' dilemmas according to Definition 10. Moving to $n \geq 2$ players, the definitions of multi-player prisoners' dilemmas, snowdrift games, and stag hunts are related (but not equal) to previous definitions of such multi-player games.

First, Definition 10.1 is related to previous definitions of a (multi-player) prisoners' dilemma (Bonacich, 1976; Taylor and Ward, 1982) and of an n-person "dilemma game" (Dawes, 1980) which require that universal C is preferred over universal D ($P_{n-1} > Q_0$, "benefits of cooperation") together with the private gains being strictly negative ($\boldsymbol{G} < \boldsymbol{0}$, "costs of cooperation"). Our definition is at the same time less and more strict than this previous definition. On the one hand, our definition is less strict in the sense that we allow for some of the private gains to be equal to zero, and hence for situations where individuals might be indifferent between one of the two choices, fixing the pure strategies of their co-players. On the other hand, our definition is more strict in the sense that action C also needs to induce positive aggregate externalities (i.e.,

$\boldsymbol{E} \gneq \boldsymbol{0}$ (40)), which previous definitions in the literature seem to have ignored. This said, in all cases a prisoners' dilemma is such that each player has no incentive to play C and that D dominates C (although only weakly, according to our definition).

Second, Definition 10.2 is related to at least one previous idea of how to generalize two-player snowdrift (a.k.a. chicken) games to more than two players. Taylor and Ward (1982) suggest that "[a] natural n-person generalization [...] is to stipulate that each player prefers to defect if 'enough' others co-operate, and to co-operate if 'too many' others defect [...] for any number of players, the preferences of any player must switch direction from 'D to C' to 'C to D' only once as the number of players choosing D increases". This is, obviously, our requirement that the sign pattern of the private gain sequence for a snowdrift game be $\varrho(\boldsymbol{G}) = (1, -1)$, i.e., that $\boldsymbol{G}$ has a single sign change from positive (incentives to cooperate when 'few' others cooperate or 'too many' others defect) to negative (incentives to defect when 'enough' others cooperate). Our definition is hence similar to this previous definition, although again stricter in the sense that we require positive aggregate externalities (40) for C to be cooperative, while Taylor and Ward (1982) only require that universal C is preferred over universal D (39).

Third, and lastly, Definition 10.3 applies a similar logic to define a (multi-player) stag hunt: here each player prefers to defect if 'few' others cooperate (or, equivalently 'too many' others defect) and prefers to cooperate if 'enough' others cooperate (or, equivalently 'few' defect), and the preferences or incentives to behave in one way or the other switch only once as the number of players choosing C (or choosing D) increases. This switch in incentives is captured by our requirement that the sign pattern of the private gain sequence for a stag hunt be $\varrho(\boldsymbol{G}) = (-1, 1)$, i.e., that $\boldsymbol{G}$ has a single sign change from negative to positive.

As a second step in our aim to generalize the notions of prisoners' dilemmas, snowdrift games and stag hunt to the multi-player case, we also introduce the following, broader definition of these games, with conditions given now in terms of the private gain function instead of the private gain sequence.

Definition 11 (Generalized prisoners' dilemmas, snowdrift games, and stag hunts)**.** Let C be cooperative, and let $\varrho(g)$ denote the sign pattern of the private gain function g.

1. We say that the game is a generalized prisoners' dilemma if $\varrho(g) = (-1)$, i.e., if $g \lneq 0$.
2. We say that the game is a generalized snowdrift game if $\varrho(g) = (1, -1)$, i.e., g has a single sign change from positive to negative.
3. We say that the game is a generalized stag hunt if $\varrho(g) = (-1, 1)$, i.e., g has a single sign change from negative to positive.

It is clear that the generalized class of each of the games in Definition 11 includes the respective class in Definition 10 (i.e., every prisoners' dilemma, snowdrift game or stag hunt is, respectively, a generalized prisoners' dilemma, a generalized snowdrift game, and a generalized stag hunt). This is because sign patterns of Bernstein transforms of sequences with at most one sign change get preserved (Lemma 3.7), which implies that, for all three games, $\varrho(g) = \varrho(\boldsymbol{G})$ holds. For $n = 2$, the converse (i.e., that every generalized prisoners' dilemma, generalized snowdrift game or generalized stag hunt is, respectively, a prisoners' dilemma, a snowdrift game, and a stag hunt) is true, because in this case the two definitions are equivalent and are simply different ways to describe the three classes of games. However, for $n > 2$ the definitions are no longer equivalent (e.g., there can be generalized stag hunts that are not "proper" stag hunts). This is due to the possible loss of sign changes when applying the Bernstein transform to a sequence $\boldsymbol{G}$ with more than one sign change (i.e., the variation-diminishing property of Bernstein transforms of Lemma 3). To illustrate, consider the teamwork dilemma introduced in Example 7, characterized by a private gain sequence with sign pattern $\varrho(\boldsymbol{G}) = (-1, 1, -1)$. In this case, it is easy to show (see Nöldeke and Peña, 2020, Lemma 1) that there exists a critical cost of contributing to the public good such that, for costs larger than such critical cost, the sign pattern of the private gain function g is $\varrho(g) = (-1)$, while for costs smaller than the critical cost the sign pattern satisfies $\varrho(g) = \varrho(\boldsymbol{G}) = (-1, 1, -1)$. In the former of these cases (large costs) the threshold public goods game with fixed costs is thus an instance of the generalized prisoners' dilemma defined in Definition 11.1. In the latter of these cases (small costs) the private gain function has more than one sign change, and the game does not fall into any of the classes covered by Definition 11.

Clearly, all games in Definition 11 are cooperative dilemmas according to our definition, as they all feature private gain functions that are negative for some interior points. Moreover, we also have the following result, which is immediate from Proposition 1.

Corollary 4. If g has at most one sign change, then a game is a cooperative dilemma if and only if it is either a generalized prisoner's dilemma, a generalized snowdrift game, or a generalized stag hunt.

Corollary 4 indicates that for *any* number of players n the generalized prisoners' dilemmas, snowdrift games, and stag hunts partition the set of cooperative dilemmas having at most one sign change in g in exactly the same way as the prisoner's dilemmas, snowdrift games, and stag hunts partition the set of cooperative dilemmas for $n = 2$. In our view this provides a sense in which these kinds of games are indeed the natural generalizations of the prototypical two-person cooperative dilemmas. Thus motivated, we proceed to take a closer look at the relationship between the ESS and social optima of these games in the following.

6.2 ESS and social optima

We now look into the ESS structure of each of the three games in Definition 11 and into the location of the social optimum $\hat{x}$ in relation to the equilibria of the game. For $n = 2$ and generic payoffs, it is the case that at an ESS individuals cooperate with a probability that is lower than what is socially optimal, i.e., $x^* < \hat{x}$ holds for all $x^* \neq \hat{x}$ (see Lemma 4). This seems intuitive: the reason why a cooperative dilemma arises is that there is a positive externality that rationality (or evolution) does not internalize. Consequently, we expect underprovision of cooperation at equilibrium. Our question is if this pattern is robust when moving to multi-player generalized prisoners' dilemmas, snowdrift games, and stag hunts.

We begin by describing the ESS structure of the games, which is an immediate consequence of Definition 11 and Lemma 1, in the following proposition.

Proposition 5 (ESS structure of generalized prisoners' dilemmas, snowdrift games, and stag hunts)**.**

1. A generalized prisoners' dilemma has exactly one ESS, namely $x^* = 0$.
2. A generalized snowdrift game has exactly one ESS $x^* \in (0, 1)$.
3. A generalized stag hunt has two ESSs: $x_1^* = 0$ and $x_2^* = 1$.

It follows from this result that all prisoners' dilemmas, snowdrift games, and stag hunts, irrespective of the number of players $n \geq 2$, feature an ESS structure like their simple two-player versions we discussed in Section 3.1. Proposition 5.2 recovers and generalizes both Gradstein and Nitzan (1990, Proposition 3) (see, also, Motro, 1991) and Anderson and Engers (2007, Proposition 1), who proved the existence and uniqueness of a symmetric NE for, respectively, the class of congestion games introduced in Example 8, and the class of public goods games with concave benefits and fixed costs introduced in Example 4. As we have seen, both of these games are particular instances of snowdrift games (and hence of generalized snowdrift games). In a similar way, Proposition 5.3 recovers and generalizes Anderson and Engers (2007, Proposition 7), who proved that $x = 0$ and $x = 1$ are symmetric NE for the class of class of games with participation synergies introduced in Example 9. Proposition 5.3 also provides a simpler proof for the result in Luo et al. (2021, Appendix A) characterizing the ASE of the replicator dynamic of their "n-person stag hunt" game (see Example 9).

We next ask whether it is the case that cooperation is underprovided at equilibrium, as it was the case for the two-player, generic versions of the games. Consider first the case of generalized prisoners' dilemmas. The following is immediate from Lemma 5.

Corollary 5. The social optimum $\hat{x}$ of every generalized prisoner's dilemma satisfies $\hat{x} > x^*$, where $x^* = 0$ is the unique ESS of the game.

In this case it is clear that the unique ESS features too little cooperation relative to the social optimum, mimicking the situation in the prisoner's dilemma with $n = 2$. Concerning the question of whether or not $\hat{x} = 1$ holds, there are (as the two-player case illustrates, see Lemma 4) generalized prisoners' dilemmas where universal C is socially optimal (and hence for which $\hat{x} = 1$ holds) and generalized prisoners' dilemmas where universal C is not socially optimal (and hence for which $0 < \hat{x} < 1$ holds).

Consider now the case of generalized snowdrift games. Here, we find again that the relation between the unique ESS and the social optimum is the same as in the underlying two-player game. More precisely, we can prove the following result.

Proposition 6. The social optimum $\hat{x}$ of every generalized snowdrift game satisfies $\hat{x} > x^*$, where $x^* \in (0, 1)$ is the unique ESS of the game.

Proof. By definition, g has a unique sign change from positive to negative and, by Proposition 5.2, a unique ESS at the totally mixed strategy $x^* \in (0, 1)$ satisfying $g(x^*) = 0$. We then have $g(x) > 0$ for all $x \in (0, x^*)$. Since C is cooperative, and by Lemma 6, h is strictly positive so that $h(x) > 0$ holds for all $x \in (0, 1)$. It then follows, via identity (29), that $f'(x) = g(x) + h(x) > 0$ holds for all $x \in (0, x^*]$. This implies that f cannot have a maximum in the interval $[0, x^*]$. Thus $\hat{x} > x^*$ must hold. □

Again, as it was the case for generalized prisoners' dilemmas, there are cases where $\hat{x} = 1$ holds (i.e., universal C is socially optimal) and cases where $0 < \hat{x} < 1$ holds (i.e., universal C is not socially optimal). Overall, Proposition 6 recovers and generalizes both Gradstein and Nitzan (1990, Proposition 7) and Anderson and Engers (2007, Proposition 2), who proved, respectively, the excessive participation at equilibrium in the class of congestion games introduced in Example 8, and the underprovision of the public good at equilibrium for the class of public goods games with concave benefits and fixed intermediate costs introduced in Example 4.

Finally, consider the case of generalized stag hunts. As in the two-player version of the game, a generalized stag hunt has exactly two ESSs, with the first at $x_1^* = 0$ and the second at $x_2^* = 1$ (Proposition 5.3). In the two-player stag hunt universal C is socially optimal, so that the social optimum $\hat{x}$ coincides with the ESS at $x_2^* = 1$. Thus, the only possibility for an inefficiency arises because $x_1^* = 0$ is an ESS featuring underprovision of cooperation.

Letting $n > 2$ opens up a new possibility, namely that universal C is no longer socially optimal (i.e., $0 < \hat{x} < 1$), so that both ESSs are inefficient, with the first of them (x_1^*) featuring "too little" cooperation, and the second (x_2^*) featuring "too much" cooperation. This possibility is illustrated in the following example.

Example 12 (continued)**.** We had seen that the game is such that C is a cooperative action, and hence satisfies $\hat{x} > 0$ by Lemma 5. Further, $\varrho(\boldsymbol{G}) = (-1, 1)$ so that the game is a stag hunt (and hence a generalized stag hunt) with $x_1^* = 0$ and $x_2^* = 1$ as the only ESSs. However, unless $\hat{x} = 1$ holds, the game has a stable rest point, namely $x_2^* = 1$, which features more cooperation than in the social optimum. The question is then if it is possible that $\hat{x} < 1$ holds. As illustrated in Fig. 1 for $z = 1/10$, this will be the case whenever $z > 0$ is sufficiently small.

Example 12 indicates that for $n > 2$ there are stag hunts which differ in a rather significant way from the two-person stag hunt. We can trace the source of this difference to the fact that in the stag hunt in Example 12 action C does not induce positive individual externalities. Indeed, when C induces positive individual externalities, we can apply Proposition 4 to obtain:

Corollary 6. Every generalized stag hunt where C induces positive individual externalities has $\hat{x} = 1$ as social optimum, so that the unique inefficient ESS $x^* = 0$ satisfies $x^* < \hat{x}$.

Proof. Every generalized stag hunt satisfies $G_{n-1} \geq 0$, as this is a necessary condition for the private gain function g to have a unique sign change from negative to positive. Thus, every generalized stag hunt belongs to the class of games for which Proposition 4 applies and yields $\hat{x} = 1$. Combining this observation with Proposition 5.3 yields the result. □

7 Concluding remarks

We have revisited the questions of what is cooperation, and what is a cooperative dilemma (Kerr et al., 2004; Nowak, 2012), in the context of binary-action multi-player games. To do so, we have mostly relied on the shape-preserving properties of Bernstein transforms. These properties have proved useful in applications ranging from approximation theory (DeVore and Lorentz, 1993) to computer-aided geometric design (Farouki, 2012). More recently, they have also been applied to game theory (Sah, 1991; Motro, 1991; Carlsson and van Damme, 1993; Menezes and Pitchford, 2006; Peña et al., 2014, 2015; Nöldeke and Peña, 2016; De Jaegher, 2019). For instance, they can be used to analyze group-size and group-size variability effects in many of the cooperative dilemmas we used to illustrate our results, and in other binary-action multi-player games (Peña and Nöldeke, 2016; Peña and Nöldeke, 2018).

We also investigated the question of whether cooperation is always underprovided at equilibrium in an inefficient equilibrium of a cooperative dilemma. To make progress, we focused on the cases of cooperative dilemmas with private gain functions having at most one sign change, i.e., the set of generalized prisoners' dilemmas, snowdrift games, and stag hunts we defined in Section 6. In doing so, we ignored other cases that can be of practical importance. A particular noteworthy example is the threshold public goods game with fixed costs and no refunds (Palfrey and Rosenthal, 1984; Bach et al., 2006) with a threshold greater than one and smaller than the group size, i.e., the game Myatt and Wallace (2008) refer to as a "teamwork dilemma", that we have characterized in Example 7. As briefly pointed out in Section 6, for sufficiently low costs, the private gain function of such a class of games has a sign pattern given by $(-1, 1, -1)$ and hence, by Lemma 1, two ESSs: $x_1^* = 0$ and $x_2^* \in (0, 1)$. It is clear that cooperation at $x_1^* = 0$ is underprovided. Is this also the case at x_2^*, i.e., is it the case that the social optimum $\hat{x}$ lies above x_2^*? Using arguments similar to the ones we used in Section 6, it can be proved that the answer to this question is positive, and that there is underprovision of cooperation at both ESSs. This result follows from two observations. First, the social gain sequence for a teamwork dilemma has the same sign pattern as its private gain sequence (namely, $(-1, 1, -1)$) and the same must be true for the social gain function. Second, it can be shown that every cooperative dilemma for which the social gain function has the sign pattern $(-1, 1, -1)$ features too little cooperation in each of its ESS. Thus, our methods can be extended to deal with more complicated scenarios. We leave this for future work.

Acknowledgements

J. Peña acknowledges funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) program, grant ANR-17-EURE-0010, and from the Institute for Advanced Study (IAS) of the University of Amsterdam.

References

Allen, B., Nowak, M.A., 2015. Games among relatives revisited. Journal of Theoretical Biology 378, 103–116. URL: https://www.sciencedirect.com/science/article/pii/S0022519315002131, doi:10.1016/j.jtbi.2015.04.031.

Anderson, S.P., Engers, M., 2007. Participation games: Market entry, coordination, and the beautiful blonde. Journal of Economic Behavior & Organization 63, 120–137. URL: https://www.sciencedirect.com/science/article/pii/S0167268106000345, doi:https://doi.org/10.1016/j.jebo.2005.05.006.

Archetti, M., Scheuring, I., 2011. Coexistence of cooperation and defection in public goods games. Evolution 65, 1140–1148. URL: https://doi.org/10.1111/j.1558-5646.2010.01185.x, doi:10.1111/j.1558-5646.2010.01185.x.

Arthur, W.B., 1994. Inductive reasoning and bounded rationality. The American Economic Review 84, 406–411. URL: http://www.jstor.org/stable/2117868.

Bach, L.A., Helvik, T., Christiansen, F.B., 2006. The evolution of n-player cooperation–threshold games and ESS bifurcations. Journal of Theoretical Biology 238, 426–434. URL: https://www.sciencedirect.com/science/article/pii/S0022519305002419, doi:10.1016/j.jtbi.2005.06.007.

Bergstrom, T., Blume, L., Varian, H., 1986. On the private provision of public goods. Journal of Public Economics 29, 25–49. URL: https://www.sciencedirect.com/science/article/pii/0047272786900241, doi:10.1016/0047-2727(86)90024-1.

Bonacich, P., S.G.H.K.J.P..M.R.J., 1976. Cooperation and group size in the n-person prisoners' dilemma. Journal of Conflict Resolution 20, 687–706.

Broom, M., Cannings, C., Vickers, G.T., 1997. Multi-player matrix games. Bulletin of Mathematical Biology 59, 931–952. URL: https://doi.org/10.1007/BF02460000, doi:10.1007/BF02460000.

Brown, L.D., Johnstone, I.M., Macgibbon, K.B., 1981. Variation diminishing transformations: A direct approach to total positivity and its statistical applications. Journal of the American Statistical Association 76, 824–832. URL: https://www.tandfonline.com/doi/abs/10.1080/01621459.1981.10477730, doi:10.1080/01621459.1981.10477730.

Buchanan, J.M., 1971. The Bases for Collective Action. General Learning Press.

Bukowski, M., Miekisz, J., 2004. Evolutionary and asymptotic stability in symmetric multi-player games. International Journal of Game Theory 33, 41–54. URL: https://doi.org/10.1007/s001820400183, doi:10.1007/s001820400183.

Carlsson, H., van Damme, E., 1993. Equilibrium selection in stag hunt games, in: K. G. Binmore, A.K., Tani, P. (Eds.), Frontiers of Game Theory. MIT Press.

Clutton-Brock, T.H., O'riain, M., Brotherton, P.N., Gaynor, D., Kansky, R., Griffin, A.S., Manser, M., 1999. Selfish sentinels in cooperative mammals. Science 284, 1640–1644.

Cohen, D., Eshel, I., 1976. On the founder effect and the evolution of altruistic traits. Theoretical Population Biology 10, 276–302.

Dawes, R.M., 1980. Social dilemmas. Annu. Rev. Psychol. 31, 169–193. URL: https://doi.org/10.1146/annurev.ps.31.020180.001125, doi:10.1146/annurev.ps.31.020180.001125.

Dawes, R.M., Orbell, J.M., Simmons, R.T., Van De Kragt, A.J.C., 1986. Organizing groups for collective action. American Political Science Review 80, 1171–1185. URL: https://www.cambridge.org/core/article/organizing-groups-for-collective-action/847B1B9E08CF2644953EC0FF9AB17EF1, doi:10.2307/1960862.

De Jaegher, K., 2017. Harsh environments and the evolution of multi-player cooperation. Theoretical Population Biology 113, 1–12. URL: https://www.sciencedirect.com/science/article/pii/S0040580916300582, doi:10.1016/j.tpb.2016.09.003.

De Jaegher, K., 2019. Harsh environments: Multi-player cooperation with excludability and congestion. Journal of Theoretical Biology 460, 18–36. URL: https://www.sciencedirect.com/science/article/pii/S0022519318304788, doi:10.1016/j.jtbi.2018.10.006.

DeVore, R.A., Lorentz, G.G., 1993. Constructive approximation. volume 303. Springer Science & Business Media.

Diekmann, A., 1985. Volunteer's dilemma. Journal of Conflict Resolution 29, 605–610.

Dindo, P., Tuinstra, J., 2011. A class of evolutionary models for participation games with negative feedback. Computational Economics 37, 267–300. URL: https://doi.org/10.1007/s10614-011-9253-3, doi:10.1007/s10614-011-9253-3.

Dixit, A., Olson, M., 2000. Does voluntary participation undermine the Coase theorem? Journal of Public Economics 76, 309–335. URL: https://www.sciencedirect.com/science/article/pii/S0047272799000894, doi:https://doi.org/10.1016/S0047-2727(99)00089-4.

Dixit, A.K., Skeath, S., McAdams, D., 2020. Games of Strategy: Fifth International Student Edition. WW Norton & Company.

Doebeli, M., Hauert, C., 2005. Models of cooperation based on the prisoner's dilemma and the snowdrift game. Ecology Letters 8, 748–766. URL: https://doi.org/10.1111/j.1461-0248.2005.00773.x, doi:10.1111/j.1461-0248.2005.00773.x.

dos Santos, M., Peña, J., 2017. Antisocial rewarding in structured populations. Scientific Reports 7, 6212. URL: https://doi.org/10.1038/s41598-017-06063-9, doi:10.1038/s41598-017-06063-9.

Farouki, R.T., 2012. The Bernstein polynomial basis: A centennial retrospective. Computer Aided Geometric Design 29, 379–419. URL: https://www.sciencedirect.com/science/article/pii/S0167839612000192, doi:10.1016/j.cagd.2012.03.001.

Fudenberg, D., Tirole, J., 1991. Game Theory. MIT press.

Gokhale, C.S., Traulsen, A., 2014. Evolutionary multiplayer games. Dynamic Games and Applications 4, 468–488. URL: https://doi.org/10.1007/s13235-014-0106-2, doi:10.1007/s13235-014-0106-2.

Gradstein, M., Nitzan, S., 1990. Binary participation and incremental provision of public goods. Social Choice and Welfare 7, 171–192. URL: https://doi.org/10.1007/BF01560583, doi:10.1007/BF01560583.

Hauert, C., Michor, F., Nowak, M.A., Doebeli, M., 2006. Synergy and discounting of cooperation in social dilemmas. Journal of Theoretical Biology 239, 195–202. URL: https://www.sciencedirect.com/science/article/pii/S0022519305003802, doi:10.1016/j.jtbi.2005.08.040.

Hilbe, C., Wu, B., Traulsen, A., Nowak, M.A., 2014. Cooperation and control in multiplayer social dilemmas. Proceedings of the National Academy of Sciences 111, 16425–16430. URL: https://doi.org/10.1073/pnas.1407887111, doi:10.1073/pnas.1407887111.

Kerr, B., Godfrey-Smith, P., Feldman, M.W., 2004. What is altruism? Trends in Ecology & Evolution 19, 135–140. URL: https://www.sciencedirect.com/science/article/pii/S0169534703003185, doi:10.1016/j.tree.2003.10.004.

Kollock, P., 1998. Social dilemmas: The anatomy of cooperation. Annual Review of Sociology 24, 183–214. URL: http://www.jstor.org/stable/223479.

Luo, Q., Liu, L., Chen, X., 2021. Evolutionary dynamics of cooperation in the n-person stag hunt game. Physica D: Nonlinear Phenomena 424, 132943. URL: https://www.sciencedirect.com/science/article/pii/S0167278921001019, doi:10.1016/j.physd.2021.132943.

Macy, M.W., Flache, A., 2002. Learning dynamics in social dilemmas. Proceedings of the National Academy of Sciences 99, 7229–7236. URL: https://doi.org/10.1073/pnas.092080099, doi:10.1073/pnas.092080099.

Makris, M., 2009. Private provision of discrete public goods. Games and Economic Behavior 67, 292–299. URL: https://www.sciencedirect.com/science/article/pii/S0899825608002030, doi:10.1016/j.geb.2008.11.003.

Matessi, C., Jayakar, S.D., 1976. Conditions for the evolution of altruism under darwinian selection. Theoretical Population Biology 9, 360–387. URL: https://www.sciencedirect.com/science/article/pii/0040580976900538, doi:10.1016/0040-5809(76)90053-8.

Matessi, C., Karlin, S., 1984. On the evolution of altruism by kin selection. Proceedings of the National Academy of Sciences 81, 1754–1758. URL: https://doi.org/10.1073/pnas.81.6.1754, doi:10.1073/pnas.81.6.1754.

Maynard Smith, J., Price, G.R., 1973. The logic of animal conflict. Nature 246, 15–18. URL: https://doi.org/10.1038/246015a0, doi:10.1038/246015a0.

McNamara, J.M., Leimar, O., 2020. Game Theory in Biology: Concepts and Frontiers. Oxford University Press.

Menezes, F.M., Pitchford, R., 2006. Binary games with many players. Economic Theory 28, 125–143. URL: https://doi.org/10.1007/s00199-005-0611-z, doi:10.1007/s00199-005-0611-z.

Motro, U., 1991. Co-operation and defection: Playing the field and the ESS. Journal of Theoretical Biology 151, 145–154. URL: https://www.sciencedirect.com/science/article/pii/S0022519305803583, doi:10.1016/S0022-5193(05)80358-3.

Myatt, D.P., Wallace, C., 2008. When does one bad apple spoil the barrel? An evolutionary analysis of collective action. The Review of Economic Studies 75, 499–527. URL: https://doi.org/10.1111/j.1467-937X.2008.00482.x, doi:10.1111/j.1467-937X.2008.00482.x.

Nowak, M.A., 2012. Evolving cooperation. Journal of Theoretical Biology 299, 1–8. URL: https://www.sciencedirect.com/science/article/pii/S002251931200015X, doi:10.1016/j.jtbi.2012.01.014.

Nöldeke, G., Peña, J., 2016. The symmetric equilibria of symmetric voter participation games with complete information. Games and Economic Behavior 99, 71–81. URL: https://www.sciencedirect.com/science/article/pii/S0899825616300641, doi:10.1016/j.geb.2016.06.016.

Nöldeke, G., Peña, J., 2020. Group size and collective action in a binary contribution game. Journal of Mathematical Economics 88, 42–51. URL: https://www.sciencedirect.com/science/article/pii/S0304406820300288, doi:10.1016/j.jmateco.2020.02.003.

Olson, M., 1965. The Logic of Collective Action: Public Goods and the Theory of Groups. Harvard University Press.

Ostrom, E., 1990. Governing the Commons: The Evolution of Institutions for Collective Action. Cambridge University Press.

Pacheco, J.M., Santos, F.C., Souza, M.O., Skyrms, B., 2009. Evolutionary dynamics of collective action in n-person stag hunt dilemmas. Proceedings of the Royal Society B: Biological Sciences 276, 315–321. URL: https://doi.org/10.1098/rspb.2008.1126, doi:10.1098/rspb.2008.1126.

Palfrey, T., Rosenthal, H., 1984. Participation and the provision of discrete public goods: a strategic analysis. Journal of Public Economics 24, 171–193.

Palfrey, T.R., Rosenthal, H., 1983. A strategic calculus of voting. Public Choice 41, 7–53. URL: https://doi.org/10.1007/BF00124048, doi:10.1007/BF00124048.

Peña, J., Nöldeke, G., 2018. Group size effects in social evolution. Journal of Theoretical Biology 457, 211–220.

Peña, J., Lehmann, L., Nöldeke, G., 2014. Gains from switching and evolutionary stability in multi-player matrix games. Journal of Theoretical Biology 346, 23–33. URL: https://www.sciencedirect.com/science/article/pii/S0022519313005675, doi:10.1016/j.jtbi.2013.12.016.

Peña, J., Nöldeke, G., 2016. Variability in group size and the evolution of collective action. Journal of Theoretical Biology 389, 72–82. URL: https://www.sciencedirect.com/science/article/pii/S0022519315005226, doi:10.1016/j.jtbi.2015.10.023.

Peña, J., Nöldeke, G., Lehmann, L., 2015. Evolutionary dynamics of collective action in spatially structured populations. Journal of Theoretical Biology 382, 122–136. URL: https://www.sciencedirect.com/science/article/pii/S0022519315003185, doi:10.1016/j.jtbi.2015.06.039.

Peña, J., Wu, B., Traulsen, A., 2016. Ordering structured populations in multiplayer cooperation games. Journal of The Royal Society Interface 13, 20150881. URL: https://doi.org/10.1098/rsif.2015.0881, doi:10.1098/rsif.2015.0881.

Płatkowski, T., 2017. On derivation and evolutionary classification of social dilemma games. Dynamic Games and Applications 7, 67–75. URL: https://doi.org/10.1007/s13235-015-0174-y, doi:10.1007/s13235-015-0174-y.

Rand, D.G., Nowak, M.A., 2013. Human cooperation. Trends in Cognitive Sciences 17, 413–425. URL: https://www.sciencedirect.com/science/article/pii/S1364661313001216, doi:10.1016/j.tics.2013.06.003.

Rapoport, A., 1987. Research paradigms and expected utility models for the provision of step-level public goods. Psychological Review 94, 74–83. doi:10.1037/0033-295X.94.1.74.

Sah, R.K., 1991. The effects of child mortality changes on fertility choice and parental welfare. Journal of Political Economy 99, 582–606. URL: https://doi.org/10.1086/261768, doi:10.1086/261768.

Santos, F.C., Pacheco, J.M., 2011. Risk of collective failure provides an escape from the tragedy of the commons. Proceedings of the National Academy of Sciences 108, 10421–10425. URL: https://doi.org/10.1073/pnas.1015648108, doi:10.1073/pnas.1015648108.

Schelling, T.C., 1973. Hockey helmets, concealed weapons, and daylight saving: A study of binary choices with externalities. Journal of Conflict Resolution 17, 381–428. URL: https://doi.org/10.1177/002200277301700302, doi:10.1177/002200277301700302.

Siegal, G., Siegal, N., Bonnie, R.J., 2009. An account of collective actions in public health. American Journal of Public Health 99, 1583–1587. URL: https://doi.org/10.2105/AJPH.2008.152629, doi:10.2105/AJPH.2008.152629.

Skyrms, B., 2004. The stag hunt and the evolution of social structure. Cambridge University Press.

Souza, M.O., Pacheco, J.M., Santos, F.C., 2009. Evolution of cooperation under n-person snowdrift games. Journal of Theoretical Biology 260, 581–588. URL: https://www.sciencedirect.com/science/article/pii/S0022519309003166, doi:10.1016/j.jtbi.2009.07.010.

Taylor, M., Ward, H., 1982. Chickens, whales, and lumpy goods: Alternative models of public-goods provision. Political Studies 30, 350–370.

Taylor, P.D., Jonker, L.B., 1978. Evolutionary stable strategies and game dynamics. Mathematical Biosciences 40, 145–156. URL: https://www.sciencedirect.com/science/article/pii/0025556478900779, doi:10.1016/0025-5564(78)90077-9.

Uyenoyama, M., Feldman, M.W., 1980. Theories of kin and group selection: A population genetics perspective. Theoretical Population Biology 17, 380–414. URL: https://www.sciencedirect.com/science/article/pii/0040580980900337, doi:10.1016/0040-5809(80)90033-7.

Van Lange, P.A.M., Joireman, J., Parks, C.D., Van Dijk, E., 2013. The psychology of social dilemmas: A review. Organizational Behavior and Human Decision Processes 120, 125–141. URL: https://www.sciencedirect.com/science/article/pii/S0749597812001276, doi:10.1016/j.obhdp.2012.11.003.

Weesie, J., Franzen, A., 1998. Cost sharing in a volunteer's dilemma. Journal of Conflict Resolution 42, 600–618. URL: https://doi.org/10.1177/0022002798042005004, doi:10.1177/0022002798042005004.

Weibull, J.W., 1995. Evolutionary Game Theory. MIT press.