Dynamic Games and Applications
https://doi.org/10.1007/s13235-023-00524-9

RESEARCH

Cooperative Dilemmas with Binary Actions and Multiple Players

Jorge Peña[1,2,3] · **Georg Nöldeke**[4]

Accepted: 3 August 2023

Abstract
The prisoner's dilemma, the snowdrift game, and the stag hunt are two-player symmetric 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 symmetric games, in particular when considering interactions among more than two players. Here, we propose definitions of these terms and explore their evolutionary consequences on the equilibrium structure of cooperative dilemmas in relation to social optimality. We show that our definition of cooperative dilemma encompasses a large class of collective action games often discussed in the literature, including congestion games, games with participation synergies, and public goods games. One of our main results is that regardless of the number of players, all cooperative dilemmas—including multi-player generalizations of the prisoner's dilemma, the snowdrift game, and the stag hunt—feature inefficient equilibria where cooperation is underprovided, but cannot have equilibria in which cooperation is overprovided. We also find simple conditions for full cooperation to be socially optimal in a cooperative dilemma. Our framework and results unify, simplify, and extend previous work on the structure and properties of cooperative dilemmas with binary actions and two or more players.

Keywords Cooperation · Social dilemmas · Cooperative dilemmas · Multiplayer games · Evolutionarily stable strategy · Bernstein transforms

This article is part of the topical collection "Evolutionary Games and Applications" edited by Christian Hilbe, Maria Kleshnina and Kateřina Staňková.

✉ Jorge Peña
jorge.pena@iast.fr; jorge_pena@eva.mpg.de

Georg Nöldeke
georg.noeldeke@unibas.ch

1 Institute for Advanced Study in Toulouse, University of Toulouse Capitole, Toulouse, France

2 Institute for Advanced Study, University of Amsterdam, Amsterdam, The Netherlands

3 Department of Human Behavior, Ecology and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

4 Faculty of Business and Economics, University of Basel, Basel, Switzerland

Published online: 26 August 2023

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 [13, 32, 35, 48, 62, 80]. 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 [6, 51], the management of common resources [52], vigilance and sentinel behavior [11], voting [55], protests, and other kinds of political participation [14], vaccination [70], mask-wearing during virus pandemics [76], and many more.

Given their ubiquity, the study of cooperative dilemmas and their resolution has attracted enormous attention in economics, political science, anthropology, psychology, sociology, and evolutionary biology. Across these different disciplines, game theory has emerged as the standard way of formalizing and thinking about cooperative dilemmas [26, 43, 82]. 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 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 [75]. 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 [42]. 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 [9].

Conceivably, the simplest game-theoretic representation of a cooperative dilemma is as a symmetric game of complete information between players who choose simultaneously between two alternative actions or strategies ("cooperation" and "defection"), i.e., a multiplayer matrix game [8, 9, 27, 57]. The most paradigmatic example of such two-strategy cooperative dilemmas is the two-player prisoner's dilemma [64]. 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) prisoner's dilemma captures 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 prisoner's dilemma, it has been realized that in many situations two other two-player games can be better representations of cooperative dilemmas occurring in nature or society: the snowdrift game [21] (or the game of chicken, [65]), and the stag hunt [71] (or assurance game). While the prisoner's 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 only by greed (but not fear) and the stag hunt is characterized only 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 prisoner's dilemma, some level of cooperation is evolutionarily stable. However, as in the prisoner's dilemma, this stable level of cooperation is lower than the socially optimal level. Second,

for the stag hunt, there are two ESSs: the first with no cooperation, 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 prisoner's dilemma, the socially optimal level of cooperation is evolutionarily stable. However, as in the prisoner's dilemma, the population can be trapped at the equilibrium where nobody cooperates. Taken together, the prisoner's dilemma, the snowdrift game, and the stag hunt constitute the three paradigmatic examples used to describe and think about cooperative dilemmas in two-player interactions [35].

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 [13, 48, 61] 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 different interpretations of the term "altruism" in evolutionary biology [34].

Here, we build on previous work [13, 33–35, 57, 59–61] 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 provide results aiming at easily identifying cooperative dilemmas, and propose multi-player generalizations of the trinity of games used in social dilemmas research, namely the prisoner's dilemma, the snowdrift game, and the stag hunt; we also show how these multi-player cooperative dilemmas have similar properties than their two-player counterparts. We consider several classes of collective action games previously discussed in the literature and show how all of these games fall into the class of cooperative dilemmas we define. We also identify simple conditions for mutual cooperation to be socially optimal, i.e., for full cooperation to maximize the expected average payoff. We end by asking whether it is the case, as it is for the two-player prisoner's dilemma, the snowdrift game, and the stag hunt, that cooperation is always underprovided at inefficient equilibria of a cooperative dilemma. A similar question has been asked before, although for more specific classes of cooperative dilemmas, by Gradstein and Nitzan [28] and Anderson and Engers [1].

2 Defining Cooperative Dilemmas

2.1 Games, Payoffs, and Strategies

We consider normal form games with two pure strategies (or actions, or choices) denoted by $\mathcal{C}$ and $\mathcal{D}$. 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. We write C_k (resp. D_k) for the payoff of a player choosing $\mathcal{C}$ (resp. $\mathcal{D}$) when k co-players choose $\mathcal{C}$. Payoffs can be

written in matrix form as

$$\begin{array}{c} \\ \mathcal{C} \\ \mathcal{D} \end{array} \begin{array}{c} \begin{array}{cccccc} n-1 & \dots & k & \dots & 1 & 0 \end{array} \\ \begin{pmatrix} C_{n-1} & \dots & C_k & \dots & C_1 & C_0 \\ D_{n-1} & \dots & D_k & \dots & D_1 & D_0 \end{pmatrix} \end{array}. \tag{1}$$

We collect the payoffs in the *payoff sequences* $\boldsymbol{C} = (C_0, C_1, \dots, C_{n-1}) \in \mathbb{R}^n$ and $\boldsymbol{D} = (D_0, D_1, \dots, D_{n-1}) \in \mathbb{R}^n$. We assume that $\boldsymbol{C} \neq \boldsymbol{D}$ holds, so as to exclude the uninteresting case where payoffs are independent of the chosen actions. However, $C_k = D_k$ may hold for some values of $k = 0, 1, \dots, n-1$, i.e., payoffs can be "non-generic." In a similar spirit, we assume that $\boldsymbol{C}$ and $\boldsymbol{D}$ 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. Throughout, the word "game" should be understood as referring to such a two-strategy symmetric game.

We consider mixed strategies represented by $x \in [0, 1]$, where x is the probability that a player chooses $\mathcal{C}$ (and $1 - x$, the probability that a player chooses $\mathcal{D}$). Thus, pure-strategy $\mathcal{D}$ (resp. $\mathcal{C}$) corresponds to mixed strategy $x = 0$ (resp. $x = 1$). We call mixed strategies with $x \in (0, 1)$, *totally mixed strategies*. Our main focus will be on symmetric strategy profiles in which all players adopt the same mixed strategy x. In particular, we will be interested in mixed strategies that are evolutionary stable when used by all players (i.e., evolutionarily stable strategies, ESS) and their relation to those strategy profiles that maximize expected average payoff over all symmetric strategy profiles. We refer to the later symmetric strategy profiles as social optima (see Sect. 3.3 for a formal definition).

2.2 What is Cooperation?

Given a game as the one described above, when can we say that action $\mathcal{C}$ corresponds to "cooperation" and action $\mathcal{D}$ to "defection"? To answer to this question, we propose the following definition of a cooperative action (hence, of "cooperation"), that we will endorse throughout.

Definition 1 (*Cooperation*) Action $\mathcal{C}$ is cooperative if and only if (i) mutual $\mathcal{C}$ is preferred over mutual $\mathcal{D}$, i.e.,

$$C_{n-1} > D_0 \tag{2}$$

holds, and (ii) action $\mathcal{C}$ induces "positive individual externalities," i.e.,

$$C_{k+1} \geq C_k \text{ and } D_{k+1} \geq D_k, \ k = 0, 1, \dots, n-2 \tag{3}$$

holds, with at least one of such inequalities being strict.

Definition 1 comprises two conditions. First, condition (i) means that players obtain a larger payoff if they all choose $\mathcal{C}$ than if they all choose $\mathcal{D}$. This condition is often encountered as part of previous definitions of multi-player "social dilemmas," "cooperative dilemmas" or "cooperation games," and hence implicitly included as a property of a cooperative action [13, 33, 48, 60, 61]. Second, condition (ii) is the requirement that the payoffs to $\mathcal{C}$-players and $\mathcal{D}$-players are non-decreasing (and at least sometimes increasing) in the number of cooperating co-players. In other words: a unilateral switch from defection to cooperation by a focal player will never decrease and will sometimes increase the payoff of a given co-player. Both stronger and weaker versions of condition (ii) of Definition 1 (and the underlying concept

of positive individual externalities) have previously appeared in the literature to characterize the meaning of a cooperative action (or cooperative type) in game-theoretic and population genetics models of multi-player cooperation. A stronger version of condition (ii) (namely that the payoff sequences $\boldsymbol{C}$ and $\boldsymbol{D}$ are both strictly increasing, i.e., Eq. (3) but with strict inequalities) appears as part of the "individual-centered" interpretation of altruism proposed by [34]. A weaker version of this condition (namely that the payoff sequences $\boldsymbol{C}$ and $\boldsymbol{D}$ are both non-decreasing without the additional requirement that one inequality holds strictly) has appeared as part of the definitions of "n-player social dilemmas" [33], "cooperation games" [60], and "multi-player social dilemmas" [61]. Our formulation of condition (ii) sits in between these previous formulations by excluding cases where payoffs are totally insensitive to the number of cooperators among co-players while allowing for cases where the payoffs to players could be, in some contexts, constant with respect to an increase in the number of cooperators. This is the case, for instance, of the volunteer's dilemma and the teamwork dilemma we discuss in Sect. 5.3.

2.3 What is a Cooperative Dilemma?

Not all games where $\mathcal{C}$ is cooperative pose a social dilemma in the sense that there is a conflict between the individual and the collective interest to choose the cooperative action. To illustrate, consider the case of a two-player game with payoff matrix

$$\begin{array}{cc} & \begin{array}{cc} \mathcal{C} & \mathcal{D} \end{array} \\ \begin{array}{c} \mathcal{C} \\ \mathcal{D} \end{array} & \begin{pmatrix} C_1 & C_0 \\ D_1 & D_0 \end{pmatrix} \end{array} \tag{4}$$

satisfying $C_1 > C_0 > D_1 > D_0$. In such a "harmony game" [37], action $\mathcal{C}$ is cooperative and the symmetric strategy profile in which both players cooperate achieves the highest possible payoff (namely C_1) for both players. Hence, $x = 1$ is the social optimum, and it is in the collective interest of the players that they both cooperate. At the same time, due to the payoff inequalities $C_0 > D_0$ and $C_1 > D_1$, cooperation is dominant. Hence, no matter what the other player does, it is in the individual interest of each player to choose $\mathcal{C}$. Individual interest and the collective interest are thus perfectly aligned and there is no discernible dilemma.

The example of the harmony game suggests to define a cooperative dilemma by adding to Definition 1 the requirement that action $\mathcal{C}$ is not dominant. This is, for instance, the approach followed by Płatkowski [61, Axiom 3]. As we will explain in Sect. 4, for games with more than two players such an approach is not satisfactory from an evolutionary perspective. Instead, to capture the intuition that a social dilemma should describe a situation where there is a conflict between the individual and the collective interests, we introduce the following definition:

Definition 2 (*Social dilemma*) A game is a social dilemma if it has an ESS x^* that is not a social optimum $\hat{x}$.

Definition 2 is similar to the one given by Kollock [35, p. 184], who defines a social dilemma as a game having "at least one deficient equilibrium." It is clear that in the harmony game the unique ESS is $x = 1$ and thus coincides with the social optimum. Hence, the harmony game is not a social dilemma according to Definition 2. The same is true for the "prisoner's delight" [72], i.e., the two-player game with payoffs satisfying $C_1 > D_1 > C_0 > D_0$, in which action $\mathcal{C}$ is cooperative and the unique ESS $x = 1$ again coincides with the social optimum.

Building on Definitions 1 and 2, we define a cooperative dilemma as follows:

Table 1 Two-player cooperative dilemmas

Name	Payoff ranking
Prisoner's dilemma	$D_1 > C_1 > D_0 \geq C_0$ or $D_1 \geq C_1 > D_0 > C_0$
Snowdrift game (or chicken)	$D_1 > C_1 \geq C_0 > D_0$
Stag hunt (or assurance game)	$C_1 > D_1 \geq D_0 > C_0$

Definition 3 (*Cooperative dilemma*) A cooperative dilemma is a social dilemma in which $\mathcal{C}$ is cooperative.

For the case of two-player games, much of the previous literature on social dilemmas has identified the prisoner's dilemma, the snowdrift game (or chicken), and the stag hunt (or assurance game) as the prototypical examples of cooperative dilemmas [35, 61]. Definition 3 is in line with this literature. Indeed, as we prove later (see Proposition 2), these are the only three different two-player games that are cooperative dilemmas according to our definition. Table 1 records the payoff inequalities that we use to define these games. Note that our definition of a prisoner's dilemma is more permissive than the standard definition, which requires $D_1 > C_1 > D_0 > C_0$. Similarly, the standard definitions of a snowdrift game and a stag hunt would require all four inequalities in the corresponding lines of Table 1 to be strict. Our definitions are in line with allowing weak inequalities in condition (ii) of Definition 1 and avoid cumbersome case distinctions.

In Sects. 4–6, we will explore the consequences of Definition 3 for the general class of games defined in Sect. 2.1. Doing so for games with more than two players is much more challenging than for the two-player case and requires additional tools and definitions. We introduce these in the following section.

3 Methods

Here, we introduce the concepts and tools that we use to derive our results. First, we briefly point out the sign terminology that we will use to describe sequences and functions. Second, we define the *private gain function* that determines the ESS structure of the game. Third, we define the *social gain function* that determines the social optimum. Fourth, we define the *external gain function*; this is the difference between the social function and the gain function. Fifth, and finally, we briefly introduce the *Bernstein transform* of a sequence. Bernstein transforms are important for our analysis because they provide a link between the sign structure of sequences determined by the payoffs of a game and the various gain functions that matter for our analysis. See Fig. 1 for an illustrating example of some of the notions we will introduce below.

3.1 Sign Patterns of Sequences and Functions

We will make frequent use of specific language to describe the sign properties of both sequences (e.g., the private, social, and external sequences defined below) and functions (e.g., the private, social, and external functions defined below). For example, for a given sequence $\boldsymbol{A} = (A_0, A_1, \ldots, A_{n-1}) \in \mathbb{R}^n$, the initial (resp. final) sign of $\boldsymbol{A}$ refers to the sign of the first (resp. final) nonzero element of $\mathbf{A}$, and the sign pattern of $\boldsymbol{A}$ refers to the sequence

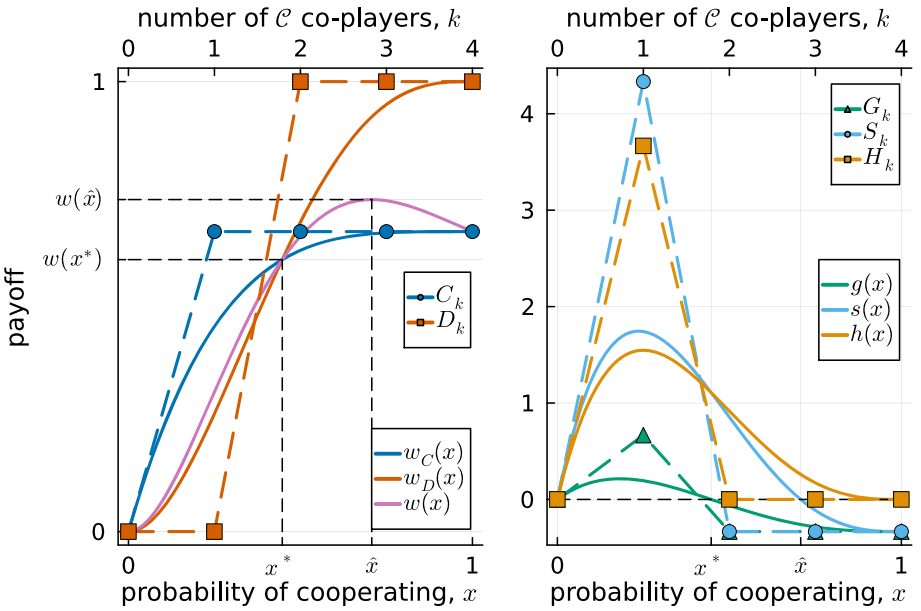

Fig. 1 Five-player game with payoff sequences $\boldsymbol{C} = (0, 2/3, 2/3, 2/3, 2/3)$ and $\boldsymbol{D} = (0, 0, 1, 1, 1)$. This is an example of the public goods games considered in Sect. 5.3 with benefit sequence given by $\boldsymbol{b} = (0, 0, 1, 1, 1, 1)$ and cost sequence given by $\boldsymbol{c} = (0, 1/3, 1/3, 1/3, 1/3)$. Action $\mathcal{C}$ is cooperative (Definition 1): (i) $C_4 = 2/3 > 0 = D_0$ holds, and (ii) $\boldsymbol{C}$ as well as $\boldsymbol{D}$ are increasing sequences. The game has a unique totally mixed ESS x^* and a unique social optimum $\hat{x}$ satisfying $0 < x^* < \hat{x} < 1$. Since $x^* \neq \hat{x}$, the game is a social dilemma (Definition 2) and, since $\mathcal{C}$ is cooperative, also a cooperative dilemma (Definition 3). Left panel: The expected payoffs to $\mathcal{C}$-players ($w_C(x)$) and to $\mathcal{D}$-players ($w_D(x)$) are, respectively, the Bernstein transforms of the payoff sequences $\boldsymbol{C}$ and $\boldsymbol{D}$. $w_C(x^*) = w_D(x^*)$ holds at the totally mixed ESS x^* (the "indifference condition" of totally mixed ESSs). The social optimum $\hat{x}$ is the global maximum of the expected average payoff $w(x)$. Right panel: The private gain function $g(x)$, the social gain function $s(x)$, and the external gain function $h(x)$ are, respectively, the Bernstein transforms of the private gain sequence $\boldsymbol{G}$, the social gain sequence $\boldsymbol{S}$, and the external gain sequence $\boldsymbol{H}$

of signs of the nonzero elements of $\mathbf{A}$ after consecutive repeated values are removed. We will also refer to a sequence as positive (resp. negative) if all of its elements are non-negative (resp. non-positive) and at least one element is positive (resp. negative). We refer the reader to "Appendix A" for more formal definitions of these and other related concepts. Similar sign notions apply to real functions; see "Appendix B" for formal definitions (restricted to polynomials on the unit interval, which are the class of real functions relevant for our analysis).

3.2 Private Gains

The expected payoff to a $\mathcal{C}$-player when all co-players play x is

$$w_C(x) = \sum_{k=0}^{n-1} \binom{n-1}{k} x^k (1-x)^{n-1-k} C_k. \tag{5}$$

Similarly, the expected payoff to a $\mathcal{D}$-player when all co-players play x is

$$w_D(x) = \sum_{k=0}^{n-1} \binom{n-1}{k} x^k (1-x)^{n-1-k} D_k. \tag{6}$$

We call the difference between these two expected payoffs the *private gain function* and denote it by g. It is given by

$$g(x) = w_C(x) - w_D(x) = \sum_{k=0}^{n-1} \binom{n-1}{k} x^k (1-x)^{n-1-k} G_k, \tag{7}$$

where

$$G_k = C_k - D_k, \quad k = 0, \ldots, n-1, \tag{8}$$

is the difference between the payoff to a $\mathcal{C}$-player interacting with k other $\mathcal{C}$-players and the payoff to a $\mathcal{D}$-player interacting with k other $\mathcal{C}$-players. Such a difference can be interpreted as the *private gain* enjoyed by a focal player who switches its action from $\mathcal{D}$ to $\mathcal{C}$ when k co-players play $\mathcal{C}$ and the remaining $n-1-k$ co-players play $\mathcal{D}$.[1] We collect the terms G_k in the *private gain sequence* $\boldsymbol{G} = \boldsymbol{C} - \boldsymbol{D} = (G_0, G_1, \ldots, G_{n-1}) \in \mathbb{R}^n$.

We are interested in the private gain function (7) because it determines the ESS structure of the game (see Lemma 3 in "Appendix C" for a formal statement). In particular, pure-strategy $x = 0$ (resp. $x = 1$) is an ESS if and only if g is negative (resp. positive) in a small neighborhood around $x = 0$ (resp. $x = 1$), while a totally mixed strategy x^* is an ESS if and only if g changes sign from positive to negative around x^*. Hence, totally mixed ESSs are roots of g (i.e., a totally mixed ESS x^* is a solution of $g(x^*) = 0$). Since the private gain function (7) is a polynomial of degree $n-1$, finding totally mixed ESSs involves solving a polynomial equation of degree $n-1$.

3.3 Social Gains

The *expected average payoff* of a player playing mixed strategy x when all co-players also play x is

$$w(x) = x w_C(x) + (1-x) w_D(x), \tag{9}$$

which can also be interpreted as the *expected population payoff* in a population where a proportion x of individuals are of type $\mathcal{C}$ and a proportion $1-x$ are of type $\mathcal{D}$. A *social optimum* is a mixed strategy

$$\hat{x} = \arg\max_{x \in [0,1]} w(x) \tag{10}$$

that maximizes w. For ease of exposition, we will assume throughout that the social optimum is unique, i.e., that the expected average payoff w has a single global maximum.

The expected average payoff can be rewritten as

$$w(x) = \sum_{i=0}^{n} \binom{n}{i} x^i (1-x)^{n-i} \frac{T_i}{n}, \tag{11}$$

[1] We have previously called such gains the "gains from switching" in [57] and the "direct gains from switching" in [59].

where

$$T_i = iC_{i-1} + (n-i)D_i,\ i = 0, 1, \ldots, n, \tag{12}$$

represents the total payoff to the n players when i players choose $\mathcal{C}$ and $n-i$ choose $\mathcal{D}$ (and where we have set $C_{-1} = D_n = 0$). We collect the elements T_i in the *total payoff sequence* $\boldsymbol{T} = (T_0, T_1, \ldots, T_n) \in \mathbb{R}^{n+1}$. The average payoff to the n players when i players choose $\mathcal{C}$ and $n-i$ choose $\mathcal{D}$ is then given by T_i/n.

Taking the derivative of the expression for the expected average payoff in (11) and simplifying, we obtain

$$s(x) = \frac{\mathrm{d}w(x)}{\mathrm{d}x} = \sum_{k=0}^{n-1} \binom{n-1}{k} x^k (1-x)^{n-1-k} S_k, \tag{13}$$

where

$$S_k = \Delta T_k = T_{k+1} - T_k,\ k = 0, \ldots, n-1, \tag{14}$$

is the first forward difference of the total payoffs. The terms (14) can be interpreted as the gains in total payoff to all players caused by a focal player that switches its action from $\mathcal{D}$ to $\mathcal{C}$ when k co-players play $\mathcal{C}$ and the remaining $n-1-k$ co-players play $\mathcal{D}$. For future reference, we collect such *social gains* (14) in the *social gain sequence* $\boldsymbol{S} = (S_0, S_1, \ldots, S_{n-1}) \in \mathbb{R}^n$, and call s the *social gain function*.

Candidate social optima correspond to either one of the two pure strategies (i.e., $\hat{x} = 0$ or $\hat{x} = 1$) or to a totally mixed strategy $\hat{x} \in (0, 1)$ satisfying the first-order condition $s(\hat{x}) = 0$. Specifically, a social optimum must be a local maximum of s. This observation leads to a link between the sign pattern of s and social optimality (see Lemma 4 of "Appendix C") that parallels the link between the sign pattern of g to evolutionary stability we discussed above.[2] In particular, a totally mixed social optimum $\hat{x}$ is a root of s. Since the social gain function (13) is a polynomial of degree $n-1$, finding totally mixed social optima involves solving a polynomial equation of degree $n-1$.

3.4 External Gains

The private gain function and social gain function both depend on the game structure given by the payoff sequences. It will be useful for our purposes to made this link explicit by expressing the social gain function as

$$s(x) = g(x) + h(x) \tag{15}$$

where

$$h(x) = x\frac{\mathrm{d}w_C(x)}{\mathrm{d}x} + (1-x)\frac{\mathrm{d}w_D(x)}{\mathrm{d}x} \tag{16}$$

$$= \sum_{k=0}^{n-1} \binom{n-1}{k} x^k (1-x)^{n-1-k} H_k, \tag{17}$$

with

$$H_k = S_k - G_k = k\Delta C_{k-1} + (n-1-k)\Delta D_k,\ k = 0, 1, \ldots, n-1, \tag{18}$$

[2] The link provided by Lemma 4 is however weaker, as conditions are necessary but not sufficient.

where $\Delta C_{k-1} = C_k - C_{k-1}$, $\Delta D_k = D_{k+1} - D_k$, and $C_{-1} = D_n = 0$. The terms (18) can be interpreted as the gains in payoff accrued to co-players when a focal player switches from playing $\mathcal{D}$ to playing $\mathcal{C}$ when k co-players play $\mathcal{C}$ and all other co-players play $\mathcal{D}$, and are thus equal to the social gains minus the private gains. We call these terms the *external gains*,[3] collect them in the *external gain sequence* $\boldsymbol{H} = (H_0, H_1, \ldots, H_{n-1}) \in \mathbb{R}^n$, and call h the *external gain function*. Equation (15) expresses the social gain function as the sum of the private gain function and the external gain function.

3.5 Bernstein Transforms

The expected payoffs to $\mathcal{C}$-players (5) and to $\mathcal{D}$-players (6) are *polynomials in Bernstein form* in the mixed strategy x with coefficients given, respectively, by the payoff sequences $\boldsymbol{C}$ and $\boldsymbol{D}$, i.e., the expected payoff to $\mathcal{C}$-players, $w_C(x)$ (resp. to $\mathcal{D}$-players, $w_D(x)$), can be understood as the *Bernstein transform* of the payoff sequence $\boldsymbol{C}$ (resp. $\boldsymbol{D}$). Likewise, the private gain function (7), the social gain function (13), and the external gain function (17), are all 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 [24, 57], including preservation of initial and final signs, preservation of positivity, and the variation-diminishing property (see "Appendix D"). These properties imply a tight link between the sign pattern of the private gain sequence $\boldsymbol{G}$ and the private gain function g, and between the social gain sequence $\boldsymbol{S}$ and the social gain function s (see Appendices A and B for our sign terminology when applied to sequences and polynomials). This link implies, via Lemmas 3 and 4 in "Appendix C," a connection between the sign pattern of the private gain sequence $\boldsymbol{G}$ and the ESS structure of the game on the one hand, and a connection between the sign pattern of the social gain sequence $\boldsymbol{S}$ and the location of the social optimum on the other hand. In many cases of interest, this allows us to identify a game as a cooperative dilemma and to extract key information about ESSs and social optima without the need for a more involved analysis (e.g., without the need for explicitly solving the polynomial equations needed to verify whether a totally mixed strategy is an ESS or a social optimum).

4 Identifying Cooperative Dilemmas

Having introduced our methods, we begin our investigation of multi-player cooperative dilemmas. We start by noting and recording two simple consequences of action $\mathcal{C}$ being cooperative. First, since mutual $\mathcal{C}$ is preferred over mutual $\mathcal{D}$ if action $\mathcal{C}$ is cooperative, it follows that the social optimum is greater than zero, i.e., that some cooperation is required to maximize the expected average payoff:

Lemma 1 *Suppose mutual $\mathcal{C}$ is preferred over universal $\mathcal{D}$, i.e., condition (2) holds. Then, $\hat{x} > 0$.*

Proof See "Appendix E." □

Second, if $\mathcal{C}$ is cooperative, then it induces positive individual externalities, which in turn implies that the external gain sequence must be positive. This implies (by the preservation of positivity property of Bernstein transforms, see Lemma 5.5) that the external gain function (17) is positive. Thus, we have:

[3] We have previously called such gains the "indirect gains from switching" in [59].

Lemma 2 *Suppose that $\mathcal{C}$ induces positive individual externalities, i.e., condition (3) holds (with at least one inequality being strict). Then, $h > 0$ holds; i.e., $h(x) \geq 0$ holds for all $x \in [0, 1]$ with the inequality being strict for all $x \in (0, 1)$.*

Lemma 2 implies (via identity (15) and, hence, $g(x) = s(x) - h(x)$) that the private gain function is strictly smaller than the social gain function for all $x \in (0, 1)$. Using Lemmas 1 and 2 together with Lemma 3 in "Appendix C" allows us to prove the following characterization result of multi-player cooperative dilemmas.

Proposition 1 *Suppose that action $\mathcal{C}$ is cooperative. Then, the following three statements are equivalent:*

(i) *The game is a cooperative dilemma.*
(ii) *There exists $x \in [0, 1]$ such that $g(x) < 0$.*
(iii) *$x = 1$ is not the only ESS of the game.*

Proof See "Appendix E." □

Proposition 1 provides two alternative necessary and sufficient conditions for a game with a cooperative action to be a cooperative dilemma. Condition (ii) is 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 gain in expected payoff) to choose $\mathcal{D}$ over $\mathcal{C}$ for at least some symmetric mixed-strategy profile played by co-players. Condition (iii) is that full cooperation ($x = 1$) is not the only ESS. This encompasses two possibilities. The first is that $x = 1$ is not an ESS (as in the prisoner's dilemma or the snowdrift game). The second is that $x = 1$ is an ESS but there exists at least one alternative ESS x^* satisfying $x^* < 1$ (as in the stag hunt). We note that the equivalence between conditions (i) and (iii) in the statement of Proposition 1 would be trivial if it were the case that $\mathcal{C}$ being cooperative implied that full cooperation is the social optimum ($\hat{x} = 1$). However, this is not so. For instance, it can be verified by direct calculation for the prisoner's dilemma and the snowdrift game that full cooperation is the social optimum for these games if and only if $2C_1 \geq C_0 + D_1$ holds. Section 6.1 provides further illustration.

Supposing that action $\mathcal{C}$ is cooperative (which is straightforward to check), condition (ii) in Proposition 1 allows us to identify a cooperative dilemma by inspecting the sign pattern of the private gain function. Although, in general, this check is simpler than explicitly identifying the ESS and the social optimum, and comparing them (as required by Definition 2), it can still be a non-trivial task requiring a numerical instead of an analytical treatment. Fortunately, in many cases of interest, a direct inspection of the sign pattern of the private gain sequence might be enough to identify a cooperative dilemma. Two-player games are one such case. Indeed, we have:

Proposition 2 (Two-player cooperative dilemmas) *Suppose that $n = 2$ and that action $\mathcal{C}$ is cooperative. Then, the game is a social dilemma (and hence a cooperative dilemma) if and only if $\boldsymbol{G}$ is not positive, i.e., if and only if*

$$G_0 < 0 \text{ or } G_1 < 0. \tag{19}$$

Further, condition (19) holds if and only if the game is a prisoner's dilemma, a snowdrift game or a stag hunt as defined in Table 1.

Proof See "Appendix E." □

Table 2 Two-player cooperative dilemmas

Name	Sign pattern of the private gain sequence
Prisoner's dilemma	$G_0 \leq 0, G_1 < 0$ or $G_0 < 0, G_1 \leq 0$
Snowdrift game (or chicken)	$G_0 > 0,\ G_1 < 0$
Stag hunt (or assurance game)	$G_0 < 0,\ G_1 > 0$

Condition (19) means that either $D_0 > C_0$ or $D_1 > C_1$ holds, i.e., that $\mathcal{C}$ does not weakly dominate $\mathcal{D}$. For two-player games, the absence of such a dominance relation is necessary and sufficient for players to have a strict ex-ante incentive to choose $\mathcal{D}$ rather than $\mathcal{C}$ for some mixed strategy of their co-player, ensuring that condition (ii) in Proposition 1 is satisfied. Table 2 records the sign patterns of the gain sequence associated with the three kinds of two-player cooperative dilemmas.

One might hope that replacing condition (19) by

$$G_k < 0, \quad \text{for some } k \in \{0, \ldots, n-1\} \tag{20}$$

yields a counterpart to Proposition 2 for multi-player games. Condition (20) again excludes the possibility that action $\mathcal{C}$ weakly dominates $\mathcal{D}$ and has been previously considered as part of the definition of a cooperative dilemma. For instance, condition (20) appears as part of the definition of "multi-player social dilemma" proposed by Płatkowski [61, Axiom 3].

It is easy to verify that condition (20) is indeed a necessary condition for a game to be a cooperative dilemma (in the sense of our definition): If condition (20) fails, then $g(x) \geq 0$ holds for all $x \in [0, 1]$, so that condition (ii) in Proposition 1 implies that the game is not a cooperative dilemma. We thus have:

Corollary 1 *Suppose that action $\mathcal{C}$ is cooperative. If the game is a cooperative dilemma, then condition (20) holds.*

Proof See "Appendix E." □

However, for $n > 2$, the absence of a weak dominance relationship between $\mathcal{C}$ and $\mathcal{D}$ (i.e., condition (20)) does not exclude the possibility that the social optimum coincides with the unique ESS. From an evolutionary perspective, it is thus not the case that condition (20) is sufficient to imply a divergence between the collective interest and the behavior induced by individual interest. The underlying reason is that there may be no ex-ante incentive to defect (i.e., the gain function g remains positive for all x) despite the existence of some ex-post incentives to defect (condition (20)). Technically, this is due to the variation-diminishing property of Bernstein transforms (see Lemma 5.5) and the possibility that the gain function has a smaller number of sign changes than the gain sequence. The following example illustrates this possibility (see also Fig. 2).

Example 1 Consider the three-player game with payoff sequences $\boldsymbol{C} = (1/2, 1, 2)$ and $\boldsymbol{D} = (-1/2, 5/4, 3/2)$ (Fig. 2). The private gain sequence is then $\boldsymbol{G} = (1, -1/4, 1/2)$. By Definition 1, action $\mathcal{C}$ is cooperative. Additionally, players have an ex-post incentive to defect when one of their co-players cooperate ($G_1 < 0$). However, the game does not constitute a cooperative dilemma, according to Definition 3, because its unique ESS coincides with the social optimum, which is given by $\hat{x} = 1$. This is because the private gain function (which simplifies to $g(x) = 1 - \frac{5}{2}x + 2x^2$) is positive in its domain, so that players have no ex-ante

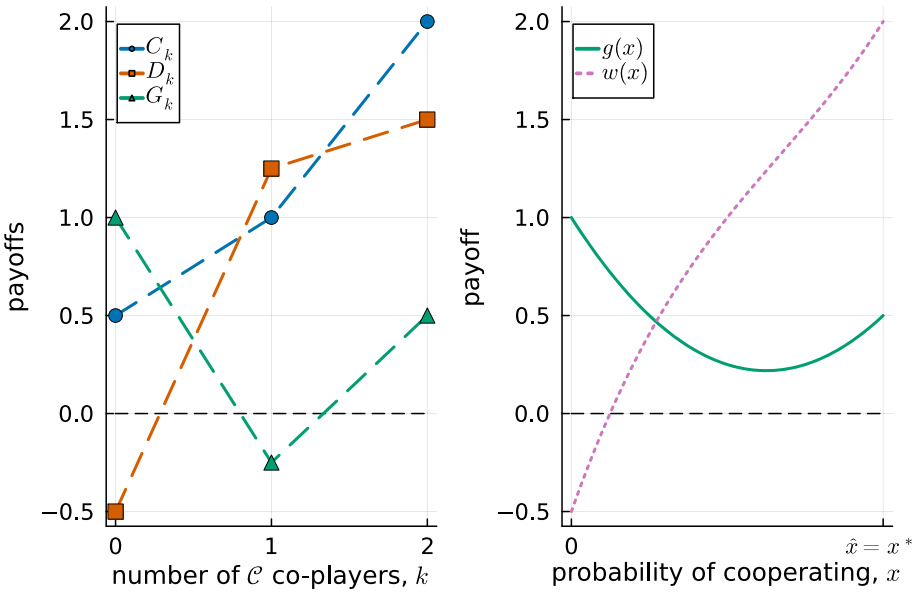

Fig. 2 Three-player game considered in Example 1. Left panel: Action $\mathcal{C}$ is cooperative (Definition 1), since (i) $C_2 > D_0$, and (ii) both the payoff sequence $\boldsymbol{C}$ and the payoff sequence $\boldsymbol{D}$ are increasing. Additionally, $G_1 < 0$ holds: individuals have an ex-post incentive to defect if exactly one co-player cooperates. Right panel: The game is not a cooperative dilemma (Definition 3), since the unique ESS $x^* = 1$ coincides with the social optimum $\hat{x} = 1$, which maximizes the expected average payoff $w(x)$. Since the private gain function $g(x)$ is never negative in the unit interval, individuals have no ex-ante incentive to defect

incentive to defect. By Proposition 1, the game is not a cooperative dilemma according to our definition.

While the generalization of (19) to (20) fails to be sufficient for a multi-player game with a cooperative action to be a cooperative dilemma, an alternative generalization of (19) yields such a sufficient condition. Specifically, consider the condition that $G_0 < 0$ or $G_{n-1} < 0$ holds, which for $n = 2$ is equivalent to (19). Due to the sign-preserving properties of Bernstein transforms, $G_0 < 0$ is equivalent to $g(0) < 0$, whereas $G_{n-1} < 0$ is equivalent to $g(1) < 0$, so that by Proposition 1, the game is a cooperative dilemma when one of these two conditions holds. More generally, it suffices that the initial sign or the final sign of the $\boldsymbol{G}$ is negative to ensure that for sufficiently small $x > 0$ or for sufficiently large $x < 1$, the inequality $g(x) < 0$ holds. Hence, we obtain:

Corollary 2 *Suppose that action $\mathcal{C}$ is cooperative. If either the initial or the final sign of $\boldsymbol{G}$ is negative, then the game is a cooperative dilemma.*

Proof See "Appendix E." □

The simplest among the multi-player games identified as cooperative dilemmas by Corollary 2 are the ones in which $\boldsymbol{G}$ is negative (ensuring that both the initial and final sign of $\boldsymbol{G}$ are negative) or the ones having only one sign change (ensuring that either the initial sign or the final sign is negative). Such games are the natural generalizations of the three kinds of two-player cooperative dilemmas to the multi-player case (cf. Table 2, which shows that for the prisoner's dilemma $\boldsymbol{G}$ is negative, whereas $\boldsymbol{G}$ has one sign change from positive to

negative for the snowdrift game and one sign change from negative to positive for the stag hunt). We are thus led to define:

Definition 4 (*Multi-player prisoner's dilemmas, snowdrift games, and stag hunts*) Let action $\mathcal{C}$ be cooperative and $n > 2$.

1. The game is a multi-player prisoner's dilemma if $\boldsymbol{G}$ is negative.
2. The game is a multi-player snowdrift game if $\boldsymbol{G}$ has a single sign change from positive to negative.
3. The game is a multi-player stag hunt if $\boldsymbol{G}$ has a single sign change from negative to positive.

Definition 4 expands the definition of prisoner's dilemmas, snowdrift games, and stag hunts to interactions among any number of players. These definitions are related (but not equal) to previous definitions of such multi-player games (for some discussion, see "Appendix F").

To see that the multi-player cooperative dilemmas defined in Definition 4 are similar to their two-player counterparts, we offer the following result, which demonstrates that the ESS structures of these games are identical to those of the corresponding two-player cooperative dilemmas.[4]

Proposition 3 (ESS structure of (multi-player) prisoner's dilemmas, snowdrift games, and stag hunts)

1. *A (multi-player) prisoner's dilemma has exactly one ESS, namely* $x^* = 0$.
2. *A (multi-player) snowdrift game has exactly one ESS* $x^* \in (0, 1)$.
3. *A (multi-player) stag hunt has two ESS, namely* $x_1^* = 0$ *and* $x_2^* = 1$.

Proof See "Appendix E." □

5 Collective Action Games as Examples of Cooperative Dilemmas

We have given a definition of a cooperative dilemma together with conditions for a multi-player game to qualify as such. In this section, we review several classes of collective action games and show how all of them represent cooperative dilemmas for suitable parameter values. Moreover, we show that many of these collective action games belong to the categories of multi-player prisoner's dilemmas, snowdrift games, and stag hunts defined in Definition 4 and characterized in Proposition 3.

5.1 Participation Games with Negative Externalities (Congestion Games)

As a first example of collective action games, consider the participation games with negative externalities to other participants (or congestion games) discussed by Anderson and Engers [1, Section 3]. This class of games includes, among others, the threshold participation game with "negative feedback" of Dindo and Tuinstra [19] and El Farol bar problem [3]; it also provides a simple formalization of the famous "tragedy of the commons" [31]. Playing $\mathcal{D}$ (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 (see Fig. 3 for an example). Playing $\mathcal{C}$ (to abstain from participating, or to stay "out") means to refrain from taking part in such

[4] Proposition 3 is not a novel result, as it is a particular case of Result 3 in [57]. We state it here for completeness.

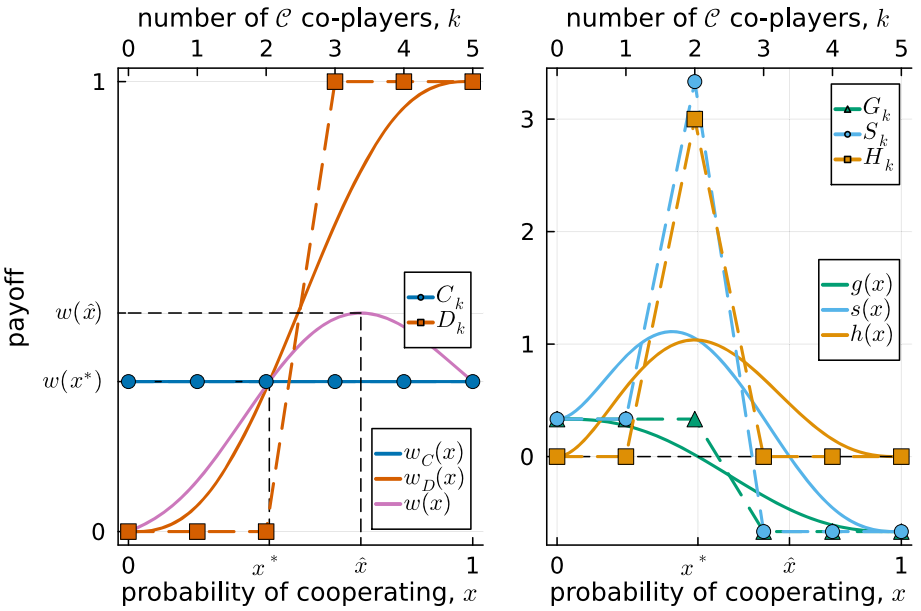

Fig. 3 Six-player congestion game with value sequence $\boldsymbol{v} = (1, 1, 1, 0, 0, 0)$ and payoff to choosing "out" of $\gamma = 1/3$. Left panel: Payoff sequences $\boldsymbol{C}$ and $\boldsymbol{D}$, payoff functions $w_C(x)$ and $w_D(x)$, and expected average payoff $w(x)$. Right panel: Private, social and external gain sequences $\boldsymbol{G}$, $\boldsymbol{S}$, and $\boldsymbol{H}$, and corresponding private, social, and external gain functions $g(x)$, $s(x)$, and $h(x)$. Action $\mathcal{C}$ is cooperative as (per Definition 1) (i) $C_5 = 1/3 > 0 = D_0$ holds, and (ii) $\boldsymbol{C}$ is constant and $\boldsymbol{D}$ is increasing. Since $\boldsymbol{G}$ has a single sign change from positive to negative, the game is classified as a multi-player snowdrift game (Definition 4.2), having a totally mixed ESS x^* (Proposition 3.2). The social optimum $\hat{x}$ features a higher probability of cooperating than the unique ESS, i.e., $x^* < \hat{x}$

an activity. The payoff to choosing "out" is a constant $\gamma > 0$ (that Anderson and Engers [1] normalize to zero). The payoff to choosing "in" is a decreasing function of the total number of $\mathcal{D}$-players. Thus, participants generate negative externalities to other participants. The payoff sequences $\boldsymbol{C}$ and $\boldsymbol{D}$ can then be written as

$$C_k = \gamma, \ k = 0, 1, \ldots, n-1, \tag{21a}$$

$$D_k = v_{n-1-k}, \ k = 0, 1, \ldots, n-1, \tag{21b}$$

where v_ℓ for $\ell \in \{0, 1, \ldots, n-1\}$ is the value of the activity to a participant ($\mathcal{D}$-player) given that ℓ co-players also participate. By assumption, the sequence $\boldsymbol{v} = (v_0, v_1, \ldots, v_{n-1}) \in \mathbb{R}^n$ is decreasing and such that $v_0 > \gamma > v_{n-1}$ holds, so that the payoff to play "in" if everybody else plays "out" is greater than the payoff to play "out," which in turn is greater than the payoff to play "in" if everybody else plays "in." It follows that $\boldsymbol{C}$ is constant and $\boldsymbol{D}$ is increasing, and hence that $\mathcal{C}$ induces positive individual externalities. Additionally, since $C_{n-1} = \gamma > v_{n-1} = D_0$ holds, action $\mathcal{C}$ (staying "out") is cooperative.

The private gains are given by

$$G_k = C_k - D_k = \gamma - v_{n-1-k}, \ k = 0, 1, \ldots, n-1. \tag{22}$$

The private gain sequence $\boldsymbol{G}$ is thus decreasing and has sign pattern $(1, -1)$, i.e., $\boldsymbol{G}$ changes sign exactly once from positive to negative. In other words, players have an incentive to participate in the activity (entering a market, exploiting a common resource, driving, going

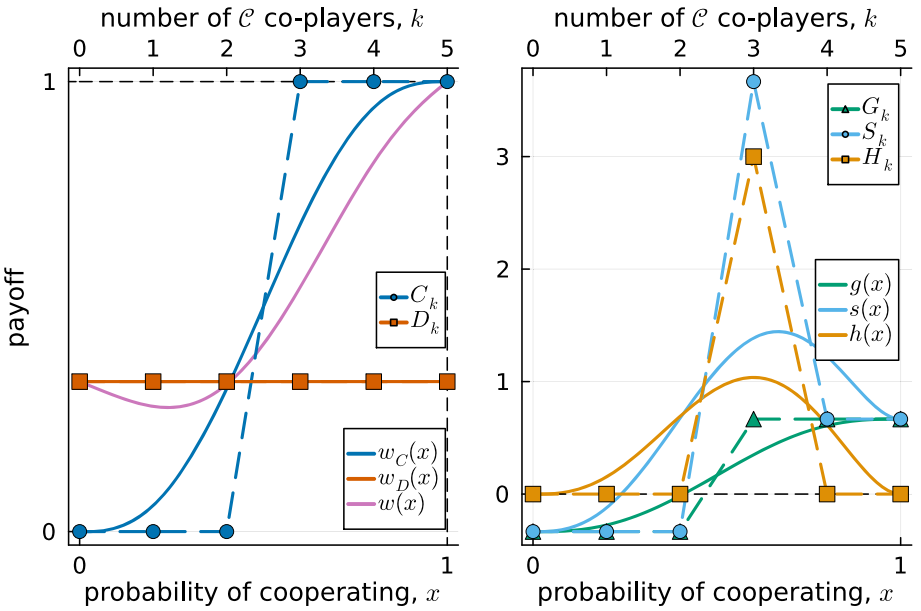

Fig. 4 Six-player game with participation synergies with value sequence $\boldsymbol{v} = (0, 0, 0, 1, 1, 1)$ and payoff to choosing "out" of $\gamma = 1/3$. Left panel: Payoff sequences $\boldsymbol{C}$ and $\boldsymbol{D}$, payoff functions $w_C(x)$ and $w_D(x)$, and expected average payoff $w(x)$. Right panel: Private, social and external gain sequences $\boldsymbol{G}$, $\boldsymbol{S}$, and $\boldsymbol{H}$, and corresponding private, social, and external gain functions $g(x)$, $s(x)$, and $h(x)$. Action $\mathcal{C}$ is cooperative as (per Definition 1) (i) $C_5 = 1 > 1/3 = D_0$ holds, and (ii) $\boldsymbol{D}$ is constant and $\boldsymbol{C}$ is increasing. Since $\boldsymbol{G}$ has a single sign change from negative to positive, the game is classified as a multi-player stag hunt (Definition 4.3). By Proposition 3.3, the game has two ESSs: $x_1^* = 0$ and $x_2^* = 1$. While $x_2^* = 1$ coincides with the social optimum, $x_1^* = 0$ is a socially inefficient ESS, where cooperation is underprovided

to a bar) as long as not too many others decide likewise. Since not participating (playing $\mathcal{C}$) is cooperative and $\boldsymbol{G}$ has a single sign change from positive to negative, congestion games are particular instances of snowdrift games (Definition 4.2). It then follows from Proposition 3.2 that congestion games are all characterized by a unique ESS x^* that is totally mixed.

5.2 Games with Participation Synergies (Strategic Complements in Participation)

As a second example of collective action games, consider the participation games with positive externalities to other participants discussed by Anderson and Engers [1, Section 4]. These games are the counterpart to those discussed in Sect. 5.1, and include the "club goods" studied by Peña et al. [59] and De Jaegher [16], and the "n-person stag hunt game" of Luo et al. [38]. Let us in this case label $\mathcal{C}$ the decision to participate, or to choose "in," and $\mathcal{D}$ 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 [1] normalize to zero). The payoff to choosing "in" is now increasing in the number of other $\mathcal{C}$-players. Thus, participants generate positive externalities to other participants (see Fig. 4 for an example). The payoff sequences $\boldsymbol{C}$ and $\boldsymbol{D}$ are given by

$$C_k = v_{k+1},\ k = 0, 1, \ldots, n-1 \tag{23a}$$

$$D_k = \gamma,\ k = 0, 1, \ldots, n-1, \tag{23b}$$

where v_i for $i \in \{0, 1, \dots, n\}$ is the value of the activity to a participant ($\mathcal{C}$-player) given the total number i of participants among players (including the self). By assumption, the sequence $\boldsymbol{v} = (v_1, \dots, v_n) \in \mathbb{R}^n$ is increasing, and such that $v_1 < \gamma < v_n$ holds, so that the payoff to play "in" if everybody else plays "out" is smaller than the payoff to play "out," which in turn is smaller than the payoff to play "in" if everybody else plays "in." It follows that $\boldsymbol{C}$ is increasing and $\boldsymbol{D}$ is constant, and hence that $\mathcal{C}$ induces positive individual externalities. Additionally, since $C_{n-1} = v_n > \gamma = D_0$ also holds, action $\mathcal{C}$ (choosing "in") is cooperative.

The private gains are given by

$$G_k = C_k - D_k = v_{k+1} - \gamma, \quad k = 0, 1, \dots, n-1. \tag{24}$$

The private gain sequence $\boldsymbol{G}$ has sign pattern $(-1, 1)$, i.e., it has a single sign change from negative to positive. In this case, players have an incentive to participate in the activity as long as sufficiently many others also decide to do so. Since participating (playing $\mathcal{C}$) is cooperative and $\boldsymbol{G}$ has a single sign change from negative to positive, games with participation synergies are particular instances of stag hunts (Definition 4.3). It then follows from Proposition 3.3 that games with participation synergies are all characterized by two ESSs: $x_1^* = 0$ and $x_2^* = 1$.

5.3 Public Goods Games

As a third and final example of collective action games, consider public goods games where playing $\mathcal{C}$ means to voluntarily contribute to a public good while playing $\mathcal{D}$ means to shirk [2, 15, 16, 20, 28, 32, 39, 53, 57, 63, 69, 73, 74, 81]. Contributing entails a cost $c_i \geq 0$ to each $\mathcal{C}$-player, while all players (both $\mathcal{C}$-players and $\mathcal{D}$-players) enjoy a benefit $b_i \geq 0$, where $0 \leq i \leq n$ denotes the total number of $\mathcal{C}$-players among players. The payoff sequences $\boldsymbol{C}$ and $\boldsymbol{D}$ are then given by

$$C_k = b_{k+1} - c_{k+1}, \ k = 0, 1, \dots, n-1 \tag{25a}$$

$$D_k = b_k, \ k = 0, 1, \dots, n-1. \tag{25b}$$

We collect the costs in the *cost sequence* $\boldsymbol{c} = (c_1, \dots, c_n) \in \mathbb{R}^n$ and the benefits in the *benefit sequence* $\boldsymbol{b} = (b_0, b_1, \dots, b_n) \in \mathbb{R}^{n+1}$. We assume that $\boldsymbol{b}$ is increasing (so that the larger the number of $\mathcal{C}$-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 $\mathcal{C}$-players never increases the cost associated with contributing). We further assume that $b_{n-1} - b_0 > c_n$ holds, so that the difference between the value of the public good if everybody contributes and its value if nobody contributes is larger than the personal cost if everybody contributes. See Figs. 1 and 5 for examples.

Since benefits $\boldsymbol{b}$ are increasing and costs $\boldsymbol{c}$ are non-decreasing, the payoff sequences $\boldsymbol{C}$ and $\boldsymbol{D}$ are increasing: Every player is better off the more other players contribute to the public good. It follows that action $\mathcal{C}$ generates positive individual externalities. Since, additionally, $b_{n-1} - b_0 > c_n$ holds, then $C_{n-1} > D_0$ holds, and action $\mathcal{C}$ is cooperative.

The private gains are given by

$$G_k = C_k - D_k = \Delta b_k - c_{k+1}, \ k = 0, 1, \dots, n-1. \tag{26}$$

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 of contributing $\Delta b_k = b_{k+1} - b_k$ scales with the number of other contributors k and compares to the cost c_{k+1}. Particular examples are given below.

Public goods games with concave benefits and fixed costs

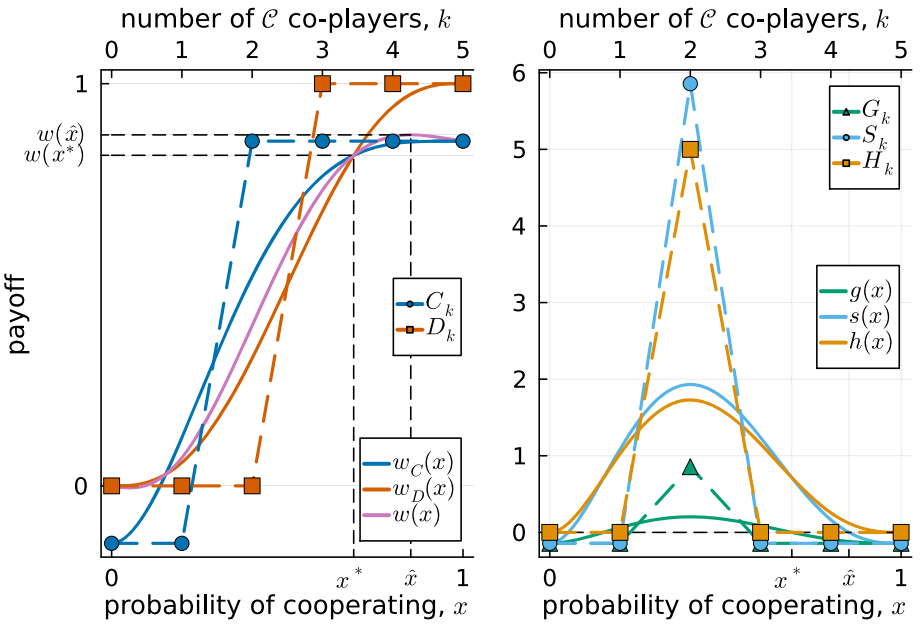

Fig. 5 Six-player public goods game with benefit sequence $\boldsymbol{b} = (0, 0, 0, 1, 1, 1)$ and cost sequence $\boldsymbol{c} = (1/7, 1/7, 1/7, 1/7, 1/7, 1/7)$, i.e., a "teamwork dilemma" with $\theta = 3$ and $\gamma = 1/7$. Left panel: Payoff sequences $\boldsymbol{C}$ and $\boldsymbol{D}$, payoff functions $w_C(x)$ and $w_D(x)$, and expected average payoff $w(x)$. Right panel: Private, social and external gain sequences $\boldsymbol{G}$, $\boldsymbol{S}$, and $\boldsymbol{H}$, and corresponding private, social, and external gain functions $g(x)$, $s(x)$, and $h(x)$. Action $\mathcal{C}$ is cooperative as (per Definition 1) (i) $C_5 = 6/7 > 0 = D_0$ holds, and (ii) both $\boldsymbol{C}$ and $\boldsymbol{D}$ are increasing. $\boldsymbol{G}$ has two sign changes: the first from negative to positive, and the second from positive to negative. The game has two ESSs: $x = 0$ and a totally mixed ESS x^*. The game also features a unique and totally mixed social optimum $\hat{x}$. Note that $0 < x^* < \hat{x}$ holds, i.e., there is underprovision of cooperation at both ESSs

Consider a public goods game where $\boldsymbol{b}$ is concave (i.e., $\Delta^2\boldsymbol{b}$ is negative) and $\boldsymbol{c}$ is a constant of value $\gamma > 0$ (i.e., $\boldsymbol{c} = (\gamma, \gamma, \ldots, \gamma)$), as assumed, for instance, by Gradstein and Nitzan [28] and Motro [45].[5] Then $\boldsymbol{G}$ is decreasing. If costs are high ($\gamma \geq \Delta b_0$), $\boldsymbol{G}$ is negative, and the game is a prisoner's dilemma (Definition 4.1). If costs are low ($\gamma \leq \Delta b_{n-1}$), $\boldsymbol{G}$ is positive, and the game is not a cooperative dilemma according to Corollary 1. 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 game is a snowdrift game (Definition 4.2). In this case, players have an individual incentive to contribute to the public good if there are relatively few contributors, and they have an incentive to shirk if there are relatively many contributors.

Public goods games with convex benefits

As a second subclass of public goods games, 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, $\boldsymbol{G}$ is increasing. If costs are high ($c_n \geq \Delta b_{n-1}$), $\boldsymbol{G}$ is negative, and the game is a prisoner's dilemma (Definition 4.1). If costs are low ($c_1 \leq \Delta b_0$), $\boldsymbol{G}$ is positive, and the game does not constitute a cooperative dilemma (Corollary 1). 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, and the game is a stag hunt (Definition 4.3).

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

Public goods games with sigmoid benefits and fixed costs

As a third subclass of public goods games, 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 studied by Pacheco et al. [53] and Archetti and Scheuring [2], where the benefit sequence first accelerates and then decelerates with the number of contributors. In this case, the private gain sequence 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, and the game is a prisoner's dilemma (Definition 4.1). If costs are low ($\gamma \leq \min_k \Delta b_k$), $\boldsymbol{G}$ is positive, and the game does not constitute a cooperative dilemma (Corollary 1). If costs are intermediate (i.e., $\min_k \Delta b_k < \gamma < \max_k \Delta b_k$ holds), then the sign pattern of $\boldsymbol{G}$ 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 $\boldsymbol{G}$ has a single sign change from positive to negative, and the game is a snowdrift game (Definition 4.2). If $\Delta b_0 < \gamma$ and $\Delta b_{n-1} \geq \gamma$, then $\boldsymbol{G}$ has a single sign change from negative to positive, and the game is a stag hunt (Definition 4.3). Finally, if $\max\{\Delta b_0, \Delta b_{n-1}\} < \gamma$, then the sign pattern of $\boldsymbol{G}$ is $(-1, 1, -1)$, i.e., the private gain sequence is first negative, then positive, and then negative again. This case, where $\boldsymbol{G}$ has two sign changes (the first one from negative to positive, the second from positive to negative) is different from the previous examples and the game cannot be classified as a prisoner's dilemma, a snowdrift or a stag hunt. Here, players have an incentive to contribute to the public good only if sufficiently many (but not too many) other players also contribute. Notwithstanding, the game is a cooperative dilemma, as it is clear by an application of Corollary 2. Moreover, the ESS structure of the game can also be characterized using Bernstein transforms [57, Results 4 and 5]. If γ is sufficiently low, the game has two ESSs: $x_1^* = 0$ and $x_2^* \in (0, 1)$; if γ is sufficiently large, the game has a unique ESS $x^* = 0$.

Threshold public goods games with fixed costs

A noteworthy example of a public goods game with sigmoid benefits and fixed costs is the threshold public goods game with fixed costs and no refunds [5, 50, 54, 74]. 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

$$b_i = [\![i \geq \theta]\!], \ i = 0, 1, \ldots, n, \tag{27}$$

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" [18]. In this case, $\boldsymbol{b}$ is concave, $\Delta b_{n-1} = 0 < \gamma < 1 = \Delta b_0$ holds, and the game is an instance of the public goods games with concave benefits and fixed intermediate costs presented in Sect. 5. In particular, the sign pattern of $\boldsymbol{G}$ is $(1, -1)$ and the game is a snowdrift game. Alternatively, if $\theta = n$ (all contributors are required), then $\boldsymbol{b}$ is convex, both $\Delta b_0 = 0 < \gamma = c_1$ and $\Delta b_{n-1} = 1 > \gamma = c_n$ hold, and the game is an instance of the public goods games with convex benefits and intermediate costs presented in Sect. 5. In particular, the sign pattern of $\boldsymbol{G}$ is $(-1, 1)$ and the game is a stag hunt. Finally, if $1 < \theta < n$ holds (more than one but less than all contributors are needed), the game is an instance of the public goods games with sigmoid benefits and fixed intermediate costs presented in Sect. 5. In this case, the game is sometimes referred to as a "teamwork dilemma" [46, 50]: the private gains are given by $G_k = -\gamma < 0$ for $k \neq \theta - 1$ and $G_{\theta-1} = 1 - \gamma > 0$, and the sign pattern of $\boldsymbol{G}$ is $(-1, 1, -1)$. 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 pivotal.

6 Cooperation and Social Optimality in Cooperative Dilemmas

In this section, we address two questions about the socially optimal probability of cooperation in a cooperative dilemma. First, we investigate whether or not the social optimum features full cooperation ($\hat{x} = 1$). While we provide a simple condition for the social optimality of full cooperation, our analysis also reveals that not all cooperative dilemmas satisfy $\hat{x} = 1$. This observation raises our second question, namely whether it could happen that a cooperative dilemma has an ESS x^* featuring overprovision of cooperation in the sense that $x^* > \hat{x}$ holds. We show that this is impossible.

6.1 When is Full Cooperation Socially Optimal?

We say that full cooperation is socially optimal if $\hat{x} = 1$ is the social optimum. It is intuitive that full cooperation should be socially optimal if, no matter which pure-strategy profile we consider, switching the action of a single player from $\mathcal{D}$ to $\mathcal{C}$ never decreases and sometimes increases the total payoff of all players. This intuition is correct. Formally, requiring that the social gain sequence $\boldsymbol{S} = (S_0, S_1, \ldots, S_{n-1})$, which we have defined in Sect. 3.3, is positive, i.e., that

$$S_k \geq 0,\ k = 0, 1, \ldots, n-1 \tag{28}$$

holds with at least one strict inequality, suffices for the optimality of full cooperation. We thus have:

Proposition 4 *Suppose $\boldsymbol{S}$ is positive. Then, the social optimum satisfies $\hat{x} = 1$.*

Proof See "Appendix G." □

To illustrate the application of Proposition 4, consider the public goods games presented in Sect. 5.3. For these games, the total payoffs are found by substituting (25) into (12) and simplifying. They are given by

$$T_i = nb_i - ic_i,\ i = 0, 1, \ldots, n, \tag{29}$$

where we set $c_0 = 0$, 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

$$S_k = \Delta T_k = n\Delta b_k - \left[(k+1)c_{k+1} - kc_k\right],\ k = 0, 1, \ldots, n-1. \tag{30}$$

Since the benefit sequence $\boldsymbol{b}$ is increasing, a sufficient condition for $\boldsymbol{S}$ to be positive is that

$$(k+1)c_{k+1} \leq kc_k,\ k = 0, 1, \ldots, n-1 \tag{31}$$

holds, i.e., that the total costs borne by contributors is non-increasing in the number of contributors. If condition (31) holds, it follows from Proposition 4 that full cooperation is socially optimal. This is the case, for instance, if there is "cost sharing" [81], i.e., if the cost sequence is given by $c_i = \gamma/i$ for some constant $\gamma > 0$.

As a counterpart to Proposition 4, we can also provide a simple sufficient condition implying that full cooperation is *not* socially optimal:

Proposition 5 *If $S_{n-1} < 0$ holds, then the social optimum satisfies $\hat{x} < 1$.*

Proof See "Appendix G." □

The condition $S_{n-1} < 0$ in the statement of Proposition 5 holds in many cooperative dilemmas. For instance, in two-player games, we have $S_{n-1} = S_1 = 2C_1 - C_0 - D_1$, so that full cooperation is not socially optimal in prisoner's dilemmas and snowdrift games with $2C_1 < C_0 + D_1$. More interestingly, Proposition 5 directly implies that the volunteer's dilemma and the teamwork dilemmas discussed in Sect. 5.3 are examples of multi-player cooperative dilemmas in which full cooperation is not socially optimal. Indeed, for such games, $\Delta b_{n-1} = 0$ (there is no additional collective benefit generated if the number of cooperators among players increases from $n-1$ to n) and $c_k = \gamma$ for all $k = 1, \ldots, n$ (costs are fixed) hold. Substituting these values into equation (30) for $k = n-1$, we obtain $S_{n-1} = -\gamma$, so that Proposition 5 applies. See also the examples illustrated in Figs. 1, 3, and 5.

6.2 Can Cooperation be Overprovided at Equilibrium?

We have made it a defining feature of a cooperative action that it induces positive individual externalities (condition (ii) in Definition 1). As these positive externalities are not internalized in the private gains that determine evolutionary stability, intuition suggests that whenever action $\mathcal{C}$ is cooperative, the probability of cooperation x^* at an ESS can not be higher than the socially optimal probability of cooperation $\hat{x}$. The following result shows that this reasoning is correct.

Proposition 6 *Suppose action $\mathcal{C}$ is cooperative. Let x^* be an ESS and $\hat{x}$ the social optimum. Then, $x^* < \hat{x}$ holds unless $x^* = \hat{x} = 1$.*

Proof See "Appendix G." □

Note that Proposition 6 not only excludes the possibility of overprovision of cooperation at an ESS ($x^* > \hat{x}$), but also establishes that the only case in which an ESS can agree with the social optimum is the one in which the internal incentives to cooperate are so strong that full cooperation is an ESS. Because the definition of a social dilemma excludes the possibility that the game has a unique ESS coinciding with the social optimum, it follows that in any cooperative dilemma either (i) all ESS feature underprovision of cooperation, $x^* < \hat{x}$ (as in the two-player prisoner's dilemma and snowdrift game) or (ii) there exist one ESS, namely $x^* = 1$, which coincides with the social optimum, but all other ESS (of which at least one exists) feature underprovision of cooperation (as in the two-player stag hunt).

We find it noteworthy and surprising that (in contrast to all other results in this paper) Proposition 6 fails if the definition of a cooperative action that we have adopted is made more permissive by replacing condition (ii) in Definition 1 with the weaker requirement that action $\mathcal{C}$ induces positive aggregate externalities, i.e., that switching one player's action from $\mathcal{D}$ to $\mathcal{C}$ never decreases but sometimes increases the sum of the co-player's payoffs. "Appendix H" expands on this.

7 Discussion

We have revisited the questions of what is cooperation, and what is a cooperative dilemma [34, 48, 61], in the context of binary-action multi-player games, and analyzed some of the

evolutionary consequences of such definitions in the absence of any additional mechanism to promote the evolution of cooperation. Our contributions are sixfold.

First, we defined an action to be cooperative if two conditions hold (Definition 1). The first condition is that mutual cooperation must provide higher payoffs than mutual defection [13]. The second condition is that cooperation must provide what we have called "positive individual externalities," that is, a player switching from defection to cooperation must never decrease and sometimes increase the payoff of each co-player for any profile of pure strategies adopted by co-players [33, 34, 60, 61, 77]. Building on this definition of a cooperative action, we then defined a cooperative dilemma as a game with a cooperative action that is also a social dilemma (Definition 3). In turn, we defined 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 average payoff (Definition 2). This definition of social dilemma is similar to the definition given by Kollock [35] but adapted to our evolutionary (and symmetric) setup. We illustrated our definitions with two-player games, and showed that the prisoner's dilemma, the snowdrift game, and the stag hunt are cooperative dilemmas according to our definition. Moreover, these three classes of games are the only kinds of cooperative dilemmas that can arise when the number of players is two.

Second, we provided simple conditions guaranteeing that a game with a cooperative action is a cooperative dilemma. A necessary and sufficient condition is that, ex-ante, players have individual incentives to defect (Proposition 1). For two-player games (and two-player games only), this condition boils down to requiring that, ex-post, players have individual incentives to defect (Proposition 2), which can be easily verified by an inspection of inequalities involving the payoffs from the game. Moving to more than two players, we provided both simple necessary (but not sufficient) and simple sufficient (but not necessary) conditions for a game to be a cooperative dilemma, given in terms of the ex-post individual incentives to defect, and hence in terms of simple inequalities involving the payoffs from the game. The necessary condition is that individuals have some ex-post incentive to defect (Corollary 1). The sufficient condition is that individuals have an ex-post incentive to defect either if everybody else defects, or if everybody else cooperates (Corollary 2).

Third, we proposed definitions of prisoner's dilemmas, snowdrift games, and stag hunts for more than two players (Definition 4). In all cases, the multi-player game has a cooperative action and an ex-post incentive structure reminiscent of its two-player counterpart, and thus stated in terms of inequalities at the level of payoffs of the game. Prisoner's dilemmas are such that defection is (weakly) dominant. Individual incentives are thus characterized by both greed (of exploiting cooperators) and fear (of being exploited by non-cooperators). 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. We showed that in all cases, the ESS structure of each of these games is reminiscent of the ESS structure of the respective two-player games (Proposition 3): the multi-player prisoner's dilemma is characterized by an unique ESS where cooperation is absent, the multi-player snowdrift game has a unique ESS that is totally mixed (so that players cooperate with a positive probability), and the stag hunt has two ESSs: full defection and full cooperation.

Fourth, we (i) reviewed three main classes of binary-action collective action games often discussed in economics and evolutionary biology, (ii) showed that they all fall (for suitable parameter values) within the category of cooperative dilemmas that we defined, and (iii) indicated how most of them fall within the class of multi-player prisoner's dilemmas, snowdrift games, and stag hunts that we defined. The first class of collective action games

comprises congestion games (Sect. 5.1), where taking part in an activity generates negative externalities to other participants, and which are special cases of snowdrift games. This observation, together with Proposition 3.2, recovers and generalizes Anderson and Engers [1, Proposition 1], who proved the existence and uniqueness of a symmetric NE for the class of congestion games we discussed. The second class comprises games with participation synergies (Sect. 5.2), where taking part in an activity generates positive externalities to other participants, and which are special cases of stag hunt games. This observation, together with Proposition 3.3, (i) recovers and strengthens Anderson and Enger [1, Proposition 7], who proved that $x = 0$ and $x = 1$ are symmetric NE for games with participation synergies, and (ii) provides a simpler proof for the result in Luo et al. [38, Appendix A] characterizing the ASE of the replicator dynamic of their "n-person stag hunt" game. The third and final class comprises the very popular public goods games (Sect. 5.3), where cooperating generates positive externalities to all other players, including individuals that do not cooperate and instead free ride on the contributions of cooperators. Here, depending on the particular shape of the benefit (or production) function, the game can be a prisoner's dilemma, a snowdrift game, a stag hunt, or a different kind of game, characterized by a private gain sequence having two sign changes: the first one from negative to positive; the second from positive to negative. Such a game (which may be called a "stagdrift game" as it combines properties of both the snowdrift game and the stag hunt) can lead to a ESS structure with two ESSs: the pure equilibrium where everybody defects, and a totally mixed equilibrium where players cooperate with a positive probability.

Fifth, we provided simple sufficient conditions for full cooperation to be socially optimal, and for the social optimum to be totally mixed. The sufficient condition for full cooperation to be socially optimal is that the social gains are positive (Proposition 4). This means that switching the action of a focal player from defection to cooperation never makes all of the players, taken as a block and including the focal, worse off and it sometimes make them better off. The sufficient condition for the social optimum to be totally mixed is that the total payoff to players when all players cooperate is smaller than the total payoff to players when all but one player cooperates (Proposition 5). This proposition illustrates the fact that for many cooperative dilemmas, the socially optimal strategy may not be full cooperation, but rather a mixed strategy where individuals defect with some positive probability. This is already the case for subclasses of two-player prisoner's dilemmas and snowdrift games, and it holds for a wide class of multi-player cooperative dilemmas as well.

Sixth, and finally, we investigated the question of whether cooperation is always underprovided at an inefficient ESS of a multi-player cooperative dilemma. We found that the answer is positive for all cooperative dilemmas as we defined them—just as it is the case for two-player cooperative dilemmas. In other words, cooperation can never be overprovided at equilibrium and it will always be underprovided unless the equilibrium coincides with the social optimum (Proposition 6). Our result recovers and generalizes to any cooperative dilemma both Gradstein and Nitzan [28, Proposition 7] and Anderson and Engers [1, Proposition 2], who proved, respectively, the underprovision at equilibrium for the class of public goods games with concave benefits and fixed intermediate costs introduced in Sect. 5, and the excessive participation at equilibrium in the class of congestion games considered in Sect. 5.1.

Previous definitions of cooperative dilemmas have proceeded axiomatically by defining cooperative dilemmas solely in terms of payoff inequalities [13, 48, 61, 62]. We find these definitions either too restrictive or too permissive. Dawes [13] defines a "social dilemma game" as a game satisfying (i) $C_{n-1} > D_0$ (i.e., our condition (2)), together with (ii) $C_k < D_k$ for all $k \in \{0, 1, \ldots, n-1\}$, i.e., the condition that the private gain sequence $\boldsymbol{G}$ is negative.

While such a definition includes the multi-player prisoner's dilemmas characterized above (and other games having $x = 0$ as the unique ESS and a social optimum satisfying $\hat{x} > 1$), many of the collective action games we have reviewed in this paper would not qualify as social dilemmas under this definition, including (multi-player) snowdrift games and stag hunts. Nowak [48, p. 2] (see also Rand and Nowak [62, Box 1]) defines a "cooperative dilemma" as a game satisfying (i) $C_{n-1} > D_0$ (i.e., our condition (2)) together with (iia) $C_k < D_k$ for some $k \in \{0, 1, \ldots, n-1\}$ or (iib) $C_{k+1} < D_k$ for some $k \in \{0, 1, \ldots, n-2\}$. Condition (iia) is equivalent to condition (20) (i.e., there is some ex-post incentive to defect, so that cooperation does not weakly dominate defection), while condition (iib) means that "in any mixed group defectors have a higher payoff than cooperators." Płatkowski [61, Definition 1] defines a "multiplayer social dilemma" as a game satisfying conditions (i) and (iia) in the definition by Nowak [48], plus requiring that the payoff sequences $\boldsymbol{C}$ and $\boldsymbol{D}$ are both non-decreasing (i.e., our condition (3) without requiring any strict inequality). The latter two definitions label as dilemmas many games where the unique ESS coincides with the unique social optimum, and hence games where there is no apparent conflict between individual and collective interests, such as the three-player game in Example 1 (see also Fig. 2). This game is both a cooperative dilemma *sensu* Nowak [48] and a multi-player social dilemma *sensu* Płatkowski [61], but not a cooperative dilemma according to our definition. By requiring that cooperative dilemmas are social dilemmas *sensu* Definition 2, and by Proposition 1, our definition of cooperative dilemma excludes such cooperative "non-dilemmas."

Our definition of cooperative action is related to the "individual-centered" definition of altruism for trait-group models in population genetics proposed by Kerr et al. [34] and based on previous work by Uyenoyama and Feldman [77]. More specifically, our condition that cooperation generates "positive individual externalities" (3) is essentially identical to conditions 7 and 8 in [34], which measure what they refer to as the "benefit of altruism."[6] Kerr et al. [34] considered two other definitions of altruism: the "focal-complement", and the "multi-level" definitions of altruism. In our framework, the way that these alternative definitions measure the benefit of altruism gives rise to potential alternative definitions of what a cooperative action is and how a cooperative dilemma could be defined (the latter by linking the new definition of cooperative action to our definition of social dilemma). Consider first the focal-complement definition of altruism proposed by Kerr et al. [34], based on previous work by Matessi and Karlin [41]. In this case, the benefit of altruism is equated with the generation of "positive aggregate externalities" (see the end of Sect. 6.2 and "Appendix H"). Indeed, our condition that cooperation generates "positive aggregate externalities" (36) is essentially identical to condition 2 in [34].[7] As we pointed out in Sect. 6.2 and "Appendix H," we could replace condition (ii) in our definition of a cooperative action for the requirement that cooperation generates positive aggregate externalities. If we make that change, then all of our results would continue to hold, but for the result on the impossibility of overprovision of cooperation at an ESS (Proposition 6). As Example 4 in "Appendix H" illustrates, if cooperation requires positive aggregate externalities instead of positive individual externalities, it is possible that cooperation is overprovided in stag hunt games.

Consider next the multi-level definition of altruism proposed by Kerr et al. [34], based on previous work by Matessi and Jayakar [40] and Cohen and Eshel [12], among others.

[6] The only difference between the formulation in Kerr et al. [34] and our condition is our use of weak instead of strict inequalities.

[7] Again, the only difference between the formulation in Kerr et al. [34] and our condition is our use of weak instead of strict inequalities.

The benefit of altruism in this definition of altruism (condition 4 in [34]) is equated with the condition that the total payoffs to players strictly increases with the number of cooperators, which is equivalent to requiring that the social gains (14) are strictly positive. Suppose we were to replace condition (ii) in our definition of a cooperative action for the requirement that the social gain sequence is positive (not necessarily strictly). Then, by Proposition 4, full cooperation would be socially optimal, and all of our results would continue to hold (except for, obviously, Proposition 5). In particular, since the social optimum features full cooperation, it would follow trivially that any ESS different from full cooperation would feature underprovision of cooperation, and that overprovision of cooperation at an inefficient ESS would be impossible.

We focused on the questions of what is cooperation, what constitutes a cooperative dilemma, and the evolutionary consequences of such definitions; not on how cooperation can be promoted, or on the various ways that different cooperative dilemmas can be resolved [4, 29, 30, 36, 47, 67, 79, 83]. However, we hope that our definitions and results will help better inform theoretical work on the mechanisms for the evolution of cooperation and conflict resolution. For instance, assortment of similar pure strategies has been often put forward as a mechanism for the evolution of cooperation [23, 25]. It is clear that for a cooperative dilemma, perfect assortment (whereby pure-strategy cooperators and pure-strategy defectors interact exclusively with individuals of the same type) will always lead to the evolution of cooperation, since the payoff for mutual cooperation is greater than the payoff for mutual defection (condition (i) in Definition 1). Under perfect assortment, an evolutionary dynamic will hence lead to a stable equilibrium of full cooperation. Such equilibrium could however be socially inefficient, as the social optimum could be totally mixed if, for example, the condition in Proposition 5 holds. As a second, related example, consider an evolutionary model in a spatially structured population with mixed strategies under so-called δ-weak selection [84], like the one in [59]. In this case, the selection gradient determining the local attractors of the evolutionary dynamics (or convergence stable trait values) is given by a linear combination of the private gain function (7) and the external gain function (17), the last term weighted by a "scaled relatedness coefficient" that depends on demographic parameters such as group size and migration rate (see Eqs. 4, 7–9 in [59]). Then, for any cooperative dilemma as we have defined them here (or more generally, any cooperative dilemma in the broad sense, see "Appendix H"), the external gain function is positive by Lemma 2. It follows that spatially structured populations and life cycles leading to positive scaled relatedness coefficients would favor the evolution of cooperation, and the more the larger the scaled relatedness. In the limit when scaled relatedness is equal to one, the selection gradient will be equal to the social gain function (via identity (15)), and the "best" convergence stable equilibrium [66, 78] will necessarily coincide with the social optimum. If the social gain function has a single local maximum, it follows that the cooperative dilemma will be in this case completely relaxed, and the conflict between individual and collective interests fully resolved.

To derive our results, we have mostly relied on the shape-preserving properties of Bernstein transforms, which have proved useful in applications ranging from approximation theory [17] to computer-aided geometric design [24], and which have also been (either implicitly or explicitly) applied to game theory [10, 16, 44, 45, 49, 57, 59, 68]. The shape-preserving properties of Bernstein transforms can also be used to analyze group-size and group-size variability effects in many of the cooperative dilemmas we characterized in this paper, and in other binary-action multi-player games [56, 58]. Overall, we hope that our framework based on Bernstein transforms will serve as a source of inspiration to further explore cooperative dilemmas, social dilemmas, and other symmetric games with binary actions and multiple players.

Acknowledgements JP 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. We thank Benjamin Allen, Péter Bayer, and two anonymous reviewers for useful comments on a previous version of this manuscript.

Author Contributions JP and GN conceived the study, performed research, and wrote the manuscript.

Funding. Open Access funding enabled and organized by Projekt DEAL. JP 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.

Availability of data and materials The Julia code used for creating the figures of this paper is publicly available on GitHub (https://github.com/jorgeapenas/CooperativeDilemmas).

Declarations

Conflict of interest The authors have no competing interests to declare.

Ethical approval. Not applicable.

Appendix A: Sign Patterns of Sequences

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

Positive and negative sequences

We say that $\boldsymbol{A}$ is non-negative, and write $\boldsymbol{A} \geq \boldsymbol{0}$, if $A_\ell \geq 0$ holds for all $\ell = 1, \ldots, m$. We say that $\boldsymbol{A}$ is positive, and write $\boldsymbol{A} \gneq \boldsymbol{0}$ if it is non-negative and nonzero, 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 $\boldsymbol{A} > \boldsymbol{0}$. Likewise, we say that $\boldsymbol{A}$ is non-positive, and write $\boldsymbol{A} \leq \boldsymbol{0}$, if $A_\ell \leq 0$ holds for all $\ell = 1, \ldots, m$. We say that $\boldsymbol{A}$ is negative, and write $\boldsymbol{A} \lneq \boldsymbol{0}$ if it is non-positive and nonzero. We say that it is strictly negative, and write $\boldsymbol{A} < \boldsymbol{0}$, 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 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 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 $I(\boldsymbol{A})$, and write

$I(\boldsymbol{A}) = \operatorname{sgn}(A_f)$, where

$$\operatorname{sgn} x = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases} \tag{32}$$

is the sign function. Thus, $I(\boldsymbol{A}) = 1$ if the initial sign is positive and $I(\boldsymbol{A}) = -1$ if the initial sign is negative. Likewise, we call the sign of the last nonzero element of $\boldsymbol{A}$ the *final sign* of $\boldsymbol{A}$, denote it by $F(\boldsymbol{A})$, and write $F(\boldsymbol{A}) = 1$ if it is positive, and $F(\boldsymbol{A}) = -1$ 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 (32) 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, $I(\boldsymbol{A}) = 1$, $F(\boldsymbol{A}) = -1$, $\sigma(\boldsymbol{A}) = 3$, and $\varrho(\boldsymbol{A}) = (1, -1, 1, -1)$.

Appendix B: Sign Patterns of Functions

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 $p \gtrsim 0$ 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 $p \lesssim 0$, 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 p 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 p 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 $I(p)$, and write $I(p) = 1$ if it is positive, and $I(p) = -1$ 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 $F(p) = 1$ if p is positive in such a neighborhood and $F(p) = -1$ if it is negative. Clearly, $I(p) = \text{sgn}\,(p(0))$ if $p(0) \neq 0$ holds. Similarly, $F(p) = \text{sgn}\,(p(1))$ 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 $I(p)$. The sign pattern describes the sign variations of the polynomial p, conveniently summarizing all the information on initial signs, final signs, and sign changes.

Appendix C: Sign Patterns of the Gain Functions and Their Relation to Evolutionary Stability and Social Optimality

Since we have assumed that $\boldsymbol{C} \neq \boldsymbol{D}$ holds, $\boldsymbol{G} \neq \boldsymbol{0}$ holds. From Eq. (7), this in turn implies $g \neq 0$, so that the initial sign $I(g)$ and final sign $F(g)$ of g are well defined. The following result, which is simply a restatement of Bukowski and Miekisz [9, Theorem 3], provides a convenient link between the sign pattern of the private gain function, and the ESS structure of the underlying multi-player game:

Lemma 3 (Sign pattern of g and evolutionary stability) *Let g be the gain function of a symmetric two-strategy n-player game, with initial sign $I(g)$ and final sign $F(g)$. Then,*

1. $x^* = 0$ *is an ESS if and only if the initial sign of g is negative, i.e., if and only if* $I(g) = -1$.
2. $x^* = 1$ *is an ESS if and only if the final sign of g is positive, i.e., if and only if* $F(g) = 1$.
3. $x^* \in (0, 1)$ *is an ESS if and only if g changes sign from positive to negative at x^*.*

In a similar spirit, we have the following partial characterization of the social optimum in relation with the sign pattern of the social gain function.

Lemma 4 (Sign pattern of s and social optimality) *Let s be the social gain function of a symmetric two-strategy n-player game, with initial sign $I(s)$ and final sign $F(s)$. Then,*

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

Appendix D: Bernstein Transforms

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* of degree m with coefficients $\boldsymbol{c}$, 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 $\boldsymbol{c} = (c_0, c_1, \ldots, c_m) \in \mathbb{R}^{m+1}$. Equation (33) can be interpreted 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 $x \in [0, 1]$.

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., [24].

Lemma 5 (Properties of Bernstein transforms) *Let $p(x) = \mathcal{B}_m(x; \boldsymbol{c})$ be a polynomial in Bernstein form of degree m with 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 $\boldsymbol{c} \neq \boldsymbol{0}$. Then, the initial and final signs of $p(x)$ and $\boldsymbol{c}$ coincide, i.e., $I(p) = I(\boldsymbol{c})$ and $F(p) = F(\boldsymbol{c})$.*
4. **Preservation of positivity** *The Bernstein transform of a positive (resp. negative) sequence is strictly positive (resp. strictly negative), i.e., if $\boldsymbol{c} \gneq \boldsymbol{0}$, then $p > 0$ (resp. if $\boldsymbol{c} \lneq \boldsymbol{0}$, 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 $j \ge 0$ is an integer.*
6. **Derivative property** *The derivative of a polynomial in Bernstein form with coefficients $\boldsymbol{c}$ is proportional to 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 pattern of p coincides with the sign pattern of $\boldsymbol{c}$. That is, if $\sigma(\boldsymbol{c}) \le 1$, then $\varrho(p) = \varrho(\boldsymbol{c})$.*

Appendix E: Omitted Proofs for Sect. 4

Proof of Lemma 1 If mutual $\mathcal{C}$ is preferred over mutual $\mathcal{D}$, then by (2) and the end-point values property of Bernstein transforms (Lemma 5.2 in "Appendix D"), $w(1) = w_C(1) = C_{n-1} > D_0 = w_D(0) = w(0)$ holds, implying that $x = 0$ does not maximize the expected average payoff w. Hence, $\hat{x} \neq 0$ and thus $\hat{x} > 0$ holds. □

Proof of Proposition 1 Let $\mathcal{C}$ be cooperative. Then, by Lemma 1, we have $\hat{x} > 0$ and, by Lemma 2, we have $h(x) > 0$ for all $x \in (0, 1)$. 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 3 in "Appendix C," $x^* = 0$ is the unique ESS. As $\hat{x} > 0$ holds, the game is a social dilemma and, therefore, a cooperative dilemma.

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

2. If g changes sign at least once (i.e., $\sigma(g) \geq 1$), there exist x such that $g(x) < 0$. Then, we have the following two subcases.
   (a) If the initial sign of g is negative (i.e., $I(g) = -1$), then, by Lemma 3, $x^* = 0$ 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., $I(g) = 1$), then, by Lemma 3, there exists at least one totally mixed ESS $x^* \in (0, 1)$. Such a totally mixed ESS satisfies the condition $g(x^*) = 0$. We then have $h(x^*) > 0$, which implies $s(x^*) > 0$ via identity (15). As x^* is totally mixed, 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 3, $x^* = 1$ is the unique ESS. In addition, since $h(x) > 0$ holds for all $x \in (0, 1)$, the identity (15) implies that $s(x) > 0$ holds for all $x \in (0, 1)$. As the social gain function s is the derivative of the average expected payoff w (see equation (13)), $\hat{x} = 1$ follows. Since the unique ESS coincides with the social optimum, the game is not a social dilemma and, therefore, not a cooperative dilemma.

Proof of Proposition 2 By Proposition 1, it suffices to check that the stated condition is equivalent to the existence of $x \in [0, 1]$ such that $g(x) < 0$ holds. If $G_0 < 0$ or $G_1 = G_{n-1} < 0$ holds, the existence of such an x is immediate from the preservation of initial and final signs of Bernstein transforms (Lemma 5.3 in "Appendix D") and the fact that g is the Bernstein transform of $\boldsymbol{G}$ (see equation (7)). On the other hand, if both $G_0 \geq 0$ and $G_1 \geq 0$ hold, then at least one of the inequalities is strict (see footnote 8) and it follows from the preservation of positivity (Lemma 5.4) that $g(x) \geq 0$ holds for all $x \in [0, 1]$. □

Proof of Corollary 1 Suppose (20) does not hold. Then, the gain sequence $\boldsymbol{G}$ is positive and it follows from the same argument as in the proof of Proposition 2 that the game is not a cooperative dilemma. Specifically, preservation of positivity (Lemma 5.4 in "Appendix D") implies that $g(x) \geq 0$ holds for all $x \in [0, 1]$. From Proposition 1, the game is not a cooperative dilemma. □

Proof of Corollary 2 If the initial sign or the final sign of $\boldsymbol{G}$ is negative, then it follows from the same argument as in the proof of Proposition 2 that the game is a cooperative dilemma. Specifically, preservation of initial and final signs (Lemma 5.3 in "Appendix D") implies the existence of $x \in [0, 1]$ such that $g(x) < 0$ holds, so that Proposition 1 implies the result. □

Proof of Proposition 3 We consider the three cases separately.

1. If $\boldsymbol{G}$ is negative, preservation of positivity (Lemma 5.4 in "Appendix D") implies that g is strictly negative. Lemma 3 then implies that $x^* = 0$ is the unique ESS of a (multi-player) prisoner's dilemma.
2. If $\boldsymbol{G}$ has a single sign change from positive to negative, the preservation of initial and final signs and the variation-diminishing property of Bernstein transforms (Lemma 5.5) implies that g has the same properties. Lemma 3 then implies that every (multi-player) snowdrift game has exactly one ESS x^* and that this ESS satisfies $0 < x^* < 1$.
3. If $\boldsymbol{G}$ has a single sign change from negative to positive, then the preservation of initial and final signs and the variation-diminishing property of Bernstein transforms (Lemma 5.5) implies that g has the same properties. Lemma 3 then implies that every (multi-player) stag hunt has two ESSs, namely $x^* = 0$ and $x^* = 1$.

Appendix F: Related Definitions of Multi-player Prisoner's Dilemmas, Snowdrift Games, and Stag Hunts

Our definitions of multi-player prisoner's dilemmas, snowdrift games, and stag hunt games (Definition 4) are related to previous definitions in the literature.

First, Definition 4.1 is related to previous definitions of a (multi-player) prisoner's dilemma [7, 74] and of an n-person "dilemma game" [13] which require that (i) mutual $\mathcal{C}$ is preferred over mutual $\mathcal{D}$ ($C_{n-1} > D_0$, condition (2)), and (ii) that the private gain sequence $\boldsymbol{G}$ is strictly negative. 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 we require action $\mathcal{C}$ to also induce positive individual externalities (i.e., condition (3)). This said, in all cases a prisoner's dilemma is such that each player has no incentive to play $\mathcal{C}$ and that $\mathcal{D}$ dominates $\mathcal{C}$ (although only weakly, according to our definition). It is also the case, by Lemma 1, that $\hat{x} > 0$.

Second, Definition 4.2 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 [74] suggest that "[a] natural n-person generalization [...] is to stipulate that each player prefers to defect if 'enough' others cooperate, and to cooperate if 'too many' others defect [...] for any number of players, the preferences of any player must switch direction from '$\mathcal{D}$ to $\mathcal{C}$' to '$\mathcal{C}$ to $\mathcal{D}$' only once as the number of players choosing $\mathcal{D}$ increases." This is, obviously, our requirement 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 individual externalities (3) for action $\mathcal{C}$ to be cooperative, while Taylor and Ward [74] only require that mutual $\mathcal{C}$ is preferred over mutual $\mathcal{D}$ (2).

Third, and lastly, Definition 4.3 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 $\mathcal{C}$ (or choosing $\mathcal{D}$) increases. This switch in incentives is captured by our requirement that $\boldsymbol{G}$ has a single sign change from negative to positive.

Appendix G: Omitted Proofs for Sect. 6

Proof of Proposition 4 The social gain function s has been defined as the derivative of the expected average payoff w in (13). Hence, it suffices to show that $s(x) > 0$ holds for all $x \in (0, 1)$ to conclude that $\hat{x} = 1$ is the social optimum. As we have noted in Sect. 3.5, s is the Bernstein transform of the social gain sequence $\boldsymbol{S}$. Since the latter has been assumed to be positive, the result follows from the preservation of positivity of Bernstein transforms (Lemma 5.4 in "Appendix D"). □

Proof of Proposition 5 By the preservation of initial and final signs (Lemma 5.3 in "Appendix D"), the Bernstein transform s of $\boldsymbol{S}$ inherits the final sign of $\boldsymbol{S}$. The condition $S_{n-1} < 0$ implies that this final sign is negative. The result then follows from Lemma 4.2 in "Appendix C." □

Proof of Proposition 6 We begin with some preliminaries: By condition (ii) in Definition 1 at least one of the sequences $\boldsymbol{C}$ and $\boldsymbol{D}$ is increasing and the other one is either increasing or constant. As we have noted in Sect. 3.5, the expected payoff w_C of a $\mathcal{C}$-player is the Bernstein transform of $\boldsymbol{C}$ and the expected payoff w_D of a $\mathcal{D}$-player is the Bernstein transform of $\boldsymbol{D}$. By the derivative property of Bernstein transforms (Lemma 5.6 in "Appendix D") and the preservation of positivity (Lemma 5.4), it follows that at least one of the functions w_C and w_D is strictly increasing and the other one is also either strictly increasing or (in case the corresponding sequence is constant, trivially) constant.

We now distinguish three cases:

1. Consider an ESS with $x^* = 1$. Then $g(1) \geq 0$ holds (Lemma 3.2 in "Appendix C"), so that, from the definition of the private gain function in (7), we have $w_C(1) \geq w_D(1)$. From the monotonicity properties of w_C and w_D noted in the preliminaries, we have $w_C(1) \geq w_C(x)$ and $w(1) \geq w_D(x)$ for all $x \in [0, 1)$ with at least one of these inequalities holding strictly. Using the definition of the expected average payoff as $w(x) = x w_C(x) + (1-x) w_D(x)$ in equation (9), it follows that $w(1) > w(x)$ holds for all $x \in [0, 1)$. Hence, $\hat{x} = 1$ is the social optimum.
2. Consider an ESS with $x^* \in (0, 1)$. We then have $g(x^*) = 0$, implying $w_C(x^*) = w_D(x^*)$. Applying the same logic as in the preceding case, it follows that $w(x^*) > w(x)$ holds for all $x \in [0, x^*)$, which in turn implies $\hat{x} \geq x^*$. To show that this last inequality must be strict, we can use the identity $s(x) = g(x) + h(x)$ (equation 15) and the ESS condition $g(x^*) = 0$ to infer that the average expected payoff is strictly increasing at x^* if $h(x^*) > 0$ holds. As the latter inequality is implied by Lemma 2, $\hat{x} > x^*$ follows.
3. Consider an ESS with $x^* = 0$. As the social optimum satisfies $\hat{x} > 0$ (see Lemma 1), the inequality $\hat{x} > x^*$ is then immediate.

□

Appendix H: Positive Aggregate Externalities

We say that action $\mathcal{C}$ generates positive aggregate externalities if the external gain sequence is positive, i.e., if $\boldsymbol{H} \gtrsim \boldsymbol{0}$, which amounts to requiring that

$$H_k \geq 0, \ k = 0, 1, \ldots, n-1, \tag{36}$$

holds with strict inequality for at least one k. In words, this is the requirement that if a focal player switches its action from $\mathcal{D}$ to $\mathcal{C}$, all co-players, taken as a block, are never worse off (and at least sometimes better off) for any pure-strategy profile that they adopt.

From the definition of external gains (18) and of positive individual externalities (3), it is clear that if action $\mathcal{C}$ induces positive individual externalities, then it also induces positive aggregate externalities. All examples of cooperative dilemmas in the main text are of this kind. However, the converse is not true: positive aggregate externalities do not necessarily imply positive individual externalities. The following two examples illustrate games for which action $\mathcal{C}$ is such that (i) it fulfills the first requirement of a cooperative action (condition (2)), (ii) it generates positive aggregate externalities, but (iii) it fails to induce positive individual externalities.[9]

[9] For yet another example related to public goods provision, see the model of "antisocial rewarding" analyzed by dos Santos and Peña [22, p. 8].

Example 2 (Competition with a superior choice)

Consider the game put forward by Menezes and Pitchford [44]. Individuals choose between two alternative choices $\mathcal{C}$ and $\mathcal{D}$, 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 the payoff sequence $\boldsymbol{C}$ is decreasing and the payoff sequence $\boldsymbol{D}$ is increasing. Since $\boldsymbol{C}$ is decreasing, action $\mathcal{C}$ does not induce positive individual externalities. Hence, $\mathcal{C}$ is not cooperative according to Definition 1. Let us now assume, as do Menezes and Pitchford [44], that $\mathcal{C}$ is "superior" in the sense that all players prefer $\mathcal{C}$ to $\mathcal{D}$ if the same number of players choose $\mathcal{C}$ or $\mathcal{D}$, e.g., bar $\mathcal{C}$ offers better music (or simply has more tables) than bar $\mathcal{D}$. This implies that $C_k > D_{n-1-k}$ holds for all $k = 0, 1, \ldots, n-1$, and, in particular, that $C_{n-1} > D_0$ holds. Hence, mutual $\mathcal{C}$ is preferred over mutual $\mathcal{D}$. Additionally, note that $\boldsymbol{H} \gtrsim \boldsymbol{0}$ can hold, provided that the switch from $\mathcal{D}$ to $\mathcal{C}$ by a focal player is such that the positive externality due to decreased competition experienced by all other $\mathcal{D}$-players compensates for the negative externality due to increased competition experienced by all other $\mathcal{C}$-players. Since $\boldsymbol{C}$ is decreasing and $\boldsymbol{D}$ is increasing by the assumption of competition, it is clear that $\Delta\boldsymbol{G} = \Delta\boldsymbol{C} - \Delta\boldsymbol{D} \lesssim \boldsymbol{0}$ holds, so that $\boldsymbol{G}$ is decreasing. Additionally, $C_0 > D_0$ (i.e., it is better to be alone at $\mathcal{C}$ rather than being with $n-1$ other players at $\mathcal{D}$) follows from the assumption that $\mathcal{C}$ is superior together with the assumption of competition, as $C_0 > D_{n-1} \geq D_{n-2} \geq \ldots \geq D_0$ holds [44]. Now assume, as do Menezes and Pitchford [44], that being alone at $\mathcal{D}$ is better than being at $\mathcal{C}$ and competing with everyone else. Then, $C_{n-1} < D_{n-1}$ 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.

Example 3 (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: $\mathcal{C}$ and $\mathcal{D}$. $\mathcal{C}$-players vote for $\mathcal{C}$ and $\mathcal{D}$-players vote for $\mathcal{D}$. The option with more votes gets selected (majority rule). Voting is costless. All players obtain a payoff of zero if the option they have chosen is not selected. $\mathcal{C}$-players (resp. D-players) obtain a payoff of $\alpha > 0$ (resp. $\beta > 0$) if their option is selected. The payoffs are then given by

$$C_k = \alpha[\![k \geq m]\!],\ k = 0, 1, \ldots, n-1, \tag{37a}$$

$$D_k = \beta[\![k \leq m]\!],\ k = 0, 1, \ldots, n-1. \tag{37b}$$

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

$$H_k = (\alpha - \beta)[\![k = m]\!],\ k = 0, 1, \ldots, n-1, \tag{38}$$

and hence $\boldsymbol{H} \gtrsim \boldsymbol{0}$ holds. Additionally, $C_{n-1} = \alpha > \beta = D_0$ holds, so mutual $\mathcal{C}$ is preferred over mutual $\mathcal{D}$. By substituting (37) into the definition of the private gains (8) we obtain

$$G_k = \alpha[\![k > m]\!] + (\alpha - \beta)[\![k = m]\!] - \beta[\![k < m]\!],\ k = 0, 1, \ldots, n-1. \tag{39}$$

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.

These two examples make us consider the possibility of allowing for a broader definition of cooperation ("cooperation in the broad sense"), by replacing condition (ii) in Definition 1

("cooperation in the strict sense") by the weaker requirement that action $\mathcal{C}$ induces positive aggregate externalities. This broader definition entails a broader definition of cooperative dilemma ("cooperative dilemma in the broad sense") that replaces the condition that action $\mathcal{C}$ is cooperative in the strict sense by the weaker requirement that action $\mathcal{C}$ is cooperative in the broad sense. Likewise, we can consider broader definitions of multi-player prisoner's dilemmas, snowdrift games, and stag hunts by letting $\mathcal{C}$ be cooperative in the broad sense in Definition 4. Examples 2 and 3 above are both examples of cooperative dilemmas in the broad (but not in the strict) sense. Example 2 is a multi-player snowdrift game in the broad sense; Example 3 is a multi-player stag hunt in the broad sense.

Importantly, all of our results in the main text, except for Proposition 6, carry over the broader definition of a cooperative dilemma. Regarding the counterpart of Proposition 6 for cooperative dilemmas in the broad sense, we have the following results.

First, consider the case of multi-player prisoner's dilemmas in the broad sense. The following is immediate from Lemma 1 and Proposition 3:

Corollary 3 *The social optimum $\hat{x}$ of every multi-player prisoner's dilemma in the broad sense satisfies $\hat{x} > x^*$, where $x^* = 0$ is the unique ESS of the game.*

Next, consider the case of multi-player snowdrift games in the broad sense. Here, we find again that the relation between the unique ESS and the social optimum is the same as for multi-player snowdrift games in the strict sense. More precisely:

Proposition 7 *The social optimum $\hat{x}$ of every multi-player snowdrift game in the broad sense satisfies $\hat{x} > x^*$, where $x^* \in (0, 1)$ is the unique ESS of the game.*

Proof Since $\varrho(\boldsymbol{G}) = (1, -1)$, the preservation of initial and final signs and the variation-diminishing property of Bernstein transforms (Lemma 5.5) implies that $\varrho(g) = (1, -1)$. Lemma 3 then implies that every (multi-player) snowdrift game in the broad sense has exactly one ESS x^* and that this ESS satisfies $0 < x^* < 1$ and $g(x^*) = 0$. We then have $g(x) > 0$ for all $x \in (0, x^*)$. Since $\mathcal{C}$ is cooperative in the broad sense, and by Lemma 2, h is strictly positive so that $h(x) > 0$ holds for all $x \in (0, 1)$. It then follows, via identity (15), that $s(x) = w'(x) = g(x) + h(x) > 0$ holds for all $x \in (0, x^*]$. This implies that w cannot have a maximum in the interval $[0, x^*]$. Thus $\hat{x} > x^*$ must hold. □

Proposition 7 proves the underprovision of cooperation at the unique ESS of the game considered by Menezes and Pitchford [44] and presented in Example 2.

Finally, consider the case of multi-player stag hunts in the broad sense. As for stag hunts in the strict sense, a stag hunt in the broad sense has two ESSs: $x_1^* = 0$ and $x_2^* = 1$ (Proposition 3.3). In the two-player stag hunt, full $\mathcal{C}$ 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. Moving to more than two players opens up a new possibility, namely that full $\mathcal{C}$ is no longer socially optimal (i.e., $0 < \hat{x} < 1$), so that both ESSs are inefficient, with the first ESS (x_1^*) featuring less cooperation than it is optimal, and the second ESS (x_2^*) featuring excessive cooperation. This possibility is illustrated in the following example.

Example 4 Consider the three-player game with payoff sequences given by $\boldsymbol{C} = (0, 1, 1+z)$ and $\boldsymbol{D} = (1, 2, 1)$, with $0 < z < 1/3$. The private gain sequence is then given by $\boldsymbol{G} = (-1, -1, z)$, the external gain sequence by $\boldsymbol{H} = (2, 0, 2z)$, and the social gain sequence by $\boldsymbol{S} = (1/3, -1/3, z)$. Since $C_2 = 1 + z > 1 = D_0$ and $\boldsymbol{H} \gtrsim \boldsymbol{0}$, the game is such that $\mathcal{C}$ is

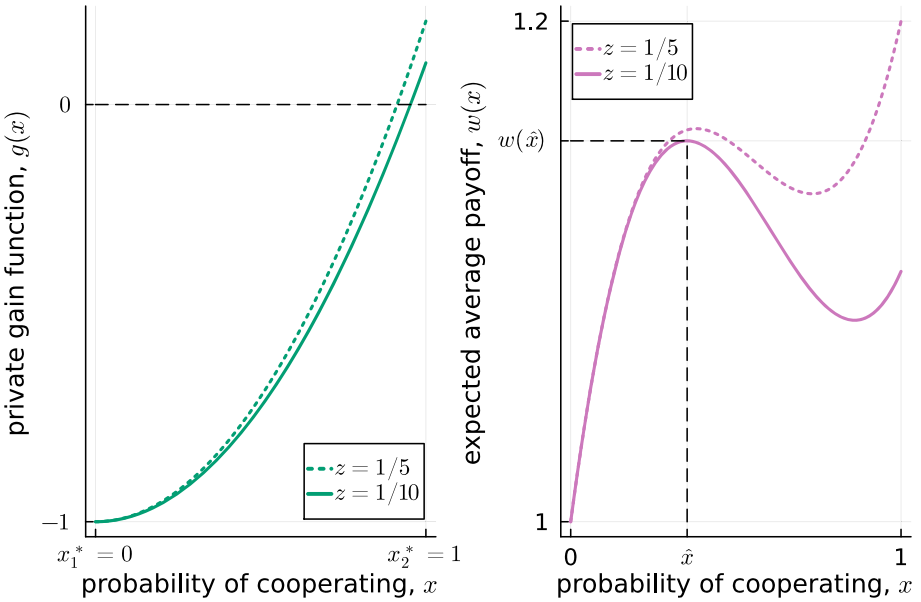

Fig. 6 Private gain functions (left panel) and expected average payoffs (right panel) for the three-player game of Example 4 for two values of z. For $z = 1/5$ (*dotted line*), the social optimum satisfies $\hat{x} = 1$ and coincides with the ESS at $x_2^* = 1$. For $z = 1/10$ (*solid line*), the social optimum satisfies $\hat{x} \approx 0.353$. In this case the social optimum is below the ESS at $x_2^* = 1$. Such an ESS then features overprovision of cooperation

cooperative in the broad sense. Further, the sign pattern of the private gain sequence is given by $\varrho(\boldsymbol{G}) = (-1, 1)$ so that the game is a stag hunt in the broad sense with $x_1^* = 0$ and $x_2^* = 1$ as ESSs. Moreover, $F(\boldsymbol{S}) = 1$ holds, so $x = 1$ locally maximizes the expected average payoff $w(x)$. However, $x = 1$ is not a global maximizer if z is sufficiently small. In this case, x_2^* features more cooperation than the social optimum, i.e., cooperation is overprovided at the ESS x_2^*. This is illustrated in Fig. 6 for $z = 1/10$, which leads to $\hat{x} \approx 0.353$.

References

1. Anderson SP, Engers M (2007) Participation games: market entry, coordination, and the beautiful blonde. J Econ Behav Org 63:120–137
2. Archetti M, Scheuring I (2011) Coexistence of cooperation and defection in public goods games. Evolution 65:1140–1148
3. Arthur WB (1994) Inductive reasoning and bounded rationality. Am Econ Rev 84:406–411
4. Axelrod R, Hamilton WD (1981) The evolution of cooperation. Science 211:1390–1396
5. Bach LA, Helvik T, Christiansen FB (2006) The evolution of n-player cooperation-threshold games and ESS bifurcations. J Theor Biol 238:426–434
6. Bergstrom T, Blume L, Varian H (1986) On the private provision of public goods. J Public Econ 29:25–49
7. Bonacich P, Shure GH, Kahan JP, Meeker RJ (1976) Cooperation and group size in the n-person prisoners' dilemma. J Conflict Resolut 20:687–706
8. Broom M, Cannings C, Vickers GT (1997) Multi-player matrix games. Bull Math Biol 59:931–952
9. Bukowski M, Miekisz J (2004) Evolutionary and asymptotic stability in symmetric multi-player games. Int J Game Theory 33:41–54
10. Carlsson H, van Damme E (1993) Equilibrium selection in stag hunt games. In: Binmore KG, Tani P (eds) Frontiers of game theory. MIT Press, Cambridge
11. Clutton-Brock TH, O'riain M, Brotherton PN, Gaynor D, Kansky R, Griffin AS, Manser M (1999) Selfish sentinels in cooperative mammals. Science 284:1640–1644
12. Cohen D, Eshel I (1976) On the founder effect and the evolution of altruistic traits. Theor Popul Biol 10:276–302
13. Dawes RM (1980) Social dilemmas. Annu Rev Psychol 31:169–193
14. Dawes RM, Orbell JM, Simmons RT, Van De Kragt AJC (1986) Organizing groups for collective action. Am Polit Sci Rev 80:1171–1185
15. De Jaegher K (2017) Harsh environments and the evolution of multi-player cooperation. Theor Popul Biol 113:1–12
16. De Jaegher K (2019) Harsh environments: Multi-player cooperation with excludability and congestion. J Theor Biol 460:18–36
17. DeVore RA, Lorentz GG (1993) Constructive approximation. Springer, New York
18. Diekmann A (1985) Volunteer's dilemma. J Confl Resolut 29:605–610
19. Dindo P, Tuinstra J (2011) A class of evolutionary models for participation games with negative feedback. Comput Econ 37:267–300
20. Dixit A, Olson M (2000) Does voluntary participation undermine the Coase theorem? J Public Econ 76:309–335
21. Doebeli M, Hauert C (2005) Models of cooperation based on the prisoner's dilemma and the snowdrift game. Ecol Lett 8:748–766
22. dos Santos M, Peña J (2017) Antisocial rewarding in structured populations. Sci Rep 7:6212
23. Eshel I, Cavalli-Sforza LL (1982) Assortment of encounters and evolution of cooperativeness. Proc Natl Acad Sci 79:1331–1335
24. Farouki RT (2012) The Bernstein polynomial basis: a centennial retrospective. Comput Aided Geom Des 29:379–419
25. Fletcher JA, Doebeli M (2008) A simple and general explanation for the evolution of altruism. Proc R Soc B Biol Sci 276:13–19
26. Fudenberg D, Tirole J (1991) Game theory. MIT Press, Cambridge
27. Gokhale CS, Traulsen A (2014) Evolutionary multiplayer games. Dyn Games Appl 4:468–488
28. Gradstein M, Nitzan S (1990) Binary participation and incremental provision of public goods. Soc Choice Welf 7:171–192
29. Hamilton WD (1964) The genetical evolution of social behaviour. I. J Theor Biol 7:1–16
30. Hamilton WD (1964) The genetical evolution of social behaviour. II. J Theor Biol 7:17–52
31. Hardin G (1968) The tragedy of the commons. Science 162:1243–1248
32. Hauert C, Michor F, Nowak MA, Doebeli M (2006) Synergy and discounting of cooperation in social dilemmas. J Theor Biol 239:195–202
33. Hilbe C, Wu B, Traulsen A, Nowak MA (2014) Cooperation and control in multiplayer social dilemmas. Proc Natl Acad Sci 111:16425–16430
34. Kerr B, Godfrey-Smith P, Feldman MW (2004) What is altruism? Trends Ecol Evol 19:135–140
35. Kollock P (1998) Social dilemmas: the anatomy of cooperation. Ann Rev Sociol 24:183–214
36. Lehmann L, Keller L (2006) The evolution of cooperation and altruism—a general framework and a classification of models. J Evol Biol 19:1365–1376

37. Licht AN (1999) Games commissions play: 2×2 games of international securities regulation. Yale J Int Law 24:61
38. Luo Q, Liu L, Chen X (2021) Evolutionary dynamics of cooperation in the n-person stag hunt game. Physica D 424:132943
39. Makris M (2009) Private provision of discrete public goods. Games Econom Behav 67:292–299
40. Matessi C, Jayakar SD (1976) Conditions for the evolution of altruism under Darwinian selection. Theor Popul Biol 9:360–387
41. Matessi C, Karlin S (1984) On the evolution of altruism by kin selection. Proc Natl Acad Sci U S A 81:1754–1758
42. Maynard Smith J, Price GR (1973) The logic of animal conflict. Nature 246:15–18
43. McNamara JM, Leimar O (2020) Game theory in biology: concepts and frontiers. Oxford University Press, Oxford
44. Menezes FM, Pitchford R (2006) Binary games with many players. Econ Theor 28:125–143
45. Motro U (1991) Co-operation and defection: playing the field and the ESS. J Theor Biol 151:145–154
46. Myatt DP, Wallace C (2008) When does one bad apple spoil the barrel? An evolutionary analysis of collective action. Rev Econ Stud 75:499–527
47. Nowak MA (2006) Five rules for the evolution of cooperation. Science 314:1560–1563
48. Nowak MA (2012) Evolving cooperation. J Theor Biol 299:1–8
49. Nöldeke G, Peña J (2016) The symmetric equilibria of symmetric voter participation games with complete information. Games Econ Behav 99:71–81
50. Nöldeke G, Peña J (2020) Group size and collective action in a binary contribution game. J Math Econ 88:42–51
51. Olson M (1965) The logic of collective action: public goods and the theory of groups. Harvard University Press, Cambridge
52. Ostrom E (1990) Governing the commons: the evolution of institutions for collective action. Cambridge University Press, Cambridge
53. Pacheco JM, Santos FC, Souza MO, Skyrms B (2009) Evolutionary dynamics of collective action in n-person stag hunt dilemmas. Proc R Soc B Biol Sci 276:315–321
54. Palfrey T, Rosenthal H (1984) Participation and the provision of discrete public goods: a strategic analysis. J Public Econ 24:171–193
55. Palfrey TR, Rosenthal H (1983) A strategic calculus of voting. Public Choice 41:7–53
56. Peña J, Nöldeke G (2018) Group size effects in social evolution. J Theor Biol 457:211–220
57. Peña J, Lehmann L, Nöldeke G (2014) Gains from switching and evolutionary stability in multi-player matrix games. J Theor Biol 346:23–33
58. Peña J, Nöldeke G (2016) Variability in group size and the evolution of collective action. J Theor Biol 389:72–82
59. Peña J, Nöldeke G, Lehmann L (2015) Evolutionary dynamics of collective action in spatially structured populations. J Theor Biol 382:122–136
60. Peña J, Wu B, Traulsen A (2016) Ordering structured populations in multiplayer cooperation games. J R Soc Interface 13:20150881
61. Płatkowski T (2017) On derivation and evolutionary classification of social dilemma games. Dyn Games Appl 7:67–75
62. Rand DG, Nowak MA (2013) Human cooperation. Trends Cogn Sci 17:413–425
63. Rapoport A (1987) Research paradigms and expected utility models for the provision of step-level public goods. Psychol Rev 94:74–83
64. Rapoport A, Chammah AM (1965) Prisoner's dilemma: a study in conflict and cooperation, vol 165. University of Michigan Press, Ann Arbor
65. Rapoport A, Chammah AM (1966) The game of chicken. Am Behav Sci 10:10–28
66. Rousset F (2004) Genetic structure and selection in subdivided populations, vol 40. Princeton University Press, Princeton
67. Sachs J, Mueller U, Wilcox T, Bull J (2004) The evolution of cooperation. Q Rev Biol 79:135–160
68. Sah RK (1991) The effects of child mortality changes on fertility choice and parental welfare. J Polit Econ 99:582–606
69. Santos FC, Pacheco JM (2011) Risk of collective failure provides an escape from the tragedy of the commons. Proc Natl Acad Sci 108:10421–10425
70. Siegal G, Siegal N, Bonnie RJ (2009) An account of collective actions in public health. Am J Public Health 99:1583–1587
71. Skyrms B (2004) The stag hunt and the evolution of social structure. Cambridge University Press, Cambridge
72. Smead R, Forber P (2013) The evolutionary dynamics of spite in finite populations. Evolution 67:698–707

73. Souza MO, Pacheco JM, Santos FC (2009) Evolution of cooperation under n-person snowdrift games. J Theor Biol 260:581–588
74. Taylor M, Ward H (1982) Chickens, whales, and lumpy goods: alternative models of public-goods provision. Polit Stud 30:350–370
75. Taylor PD, Jonker LB (1978) Evolutionary stable strategies and game dynamics. Math Biosci 40:145–156
76. Traulsen A, Levin SA, Saad-Roy CM (2023) Individual costs and societal benefits of interventions during the COVID-19 pandemic. Proc Natl Acad Sci 120:e2303546120
77. Uyenoyama M, Feldman MW (1980) Theories of kin and group selection: a population genetics perspective. Theor Popul Biol 17:380–414
78. Van Cleve J (2015) Social evolution and genetic interactions in the short and long term. Theor Popul Biol 103:2–26
79. Van Cleve J, Akçay E (2014) Pathways to social evolution: reciprocity, relatedness, and synergy. Evolution 68:2245–2258
80. Van Lange PAM, Joireman J, Parks CD, Van Dijk E (2013) The psychology of social dilemmas: a review. Organ Behav Hum Decis Process 120:125–141
81. Weesie J, Franzen A (1998) Cost sharing in a Volunteer's dilemma. J Conflict Resolut 42:600–618
82. Weibull JW (1995) Evolutionary game theory. MIT Press, Cambridge
83. West SA, Griffin AS, Gardner A (2007) Evolutionary explanations for cooperation. Curr Biol 17:R661–R672
84. Wild G, Traulsen A (2007) The different limits of weak selection and the evolutionary dynamics of finite populations. J Theor Biol 247:382–390