CrossMark

On Derivation and Evolutionary Classification of Social Dilemma Games

Tadeusz Płatkowski[1]

Abstract We propose an axiomatic derivation and a classification of multiperson social dilemma games. For the two-person symmetric games, the axiomatization leads to three types of the social dilemmas only: the Prisoner's Dilemma (PD) game, the Chicken (Snowdrift) game and the Stag Hunt game. The popular multiperson games, such as for example the N-person PD, the Public Goods, the Tragedy of the Commons, the Volunteer's Dilemma and the Assurance game, are included in the proposed frame. For general social dilemma games with arbitrary payoffs, their simple classification is proposed, based on the number of stable equilibria of the corresponding replicator dynamics.

Keywords Social dilemmas · Multiperson games · Cooperation · Replicator dynamics

1 Introduction

Social dilemmas describe situations in which selfish decisions by autonomous individuals jointly result in group inefficient outcomes. The notion dilemma comes from the fact that acting in one's self-interest is tempting to every individual involved, even though all individuals would benefit from acting in the group interest. Social dilemmas have attracted a great deal of interest of scholars in various branches of science, among others in psychology, sociology, economy, politics, biology, cf. e.g., [5,9–11,13,16,19,21,22] and references cited therein.

The formal approach to social dilemmas is provided by game theory. The corresponding models, i.e., social dilemma games, were studied by many scholars, cf. e.g., [1,4,6,12,14,15]. In the context of N-person strategic games with two strategies (N-person binary games): cooperate and defect, Dawes, cf. [4] and references therein, characterizes social dilemma games as such in which "1. the payoff when m other people cooperate is always higher for an individual who remains a defector than for one who becomes the m plus first cooperator and 2. universal

✉ Tadeusz Płatkowski
tplatk@mimuw.edu.pl

[1] Faculty of Mathematics, Informatics, and Mechanics, University of Warsaw, Warsaw, Poland

cooperation among the N players leads to a greater payoff than does universal defection..".
He also analyzes two games (the "Take Some" and the "Give Some" games) which in addition
satisfy inequalities equivalent to strong monotonicity of the payoff functions.

Liebrand [12] defines social dilemma as a situation in which "(1) there is a strategy that
yields the person the best payoff in at least one configuration of strategy choices and that has
a negative impact on the interests of the other persons involved, and (2) the choice of that
particular strategy by all persons results in a deficient outcome." The definition of Liebrand
extended the set of social dilemmas from the PD game, studied e.g., in the social sciences
and in the evolutionary game theory as a paradigm for the evolution of cooperation, to other
two-player games: the Chicken (Snowdrift) game and the Stag Hunt (Assurance) game. These
different types of social dilemmas can be characterized by a different level of "social tension."
In particular, while the Assurance game has fear built in, the Chicken game has greed rather
then fear, whereas the PD has both fear and greed built in (formalization of these notions for
general multiperson games is presented later, cf. Sect. 3).

In [15], the authors characterize the N-person binary games through their properties,
including the number of steady states, as well as the asymptotic properties of the state
trajectories, and show, by the agent-based simulations, that at least three different types of
solutions can be obtained for the PD game by properly changing the payoff functions. They
considered a discrete dynamic system and linear payoff functions with a set of four standard
PD parameters.

In the first part of the present paper, we propose an axiomatization of the social dilemmas
for multiperson games, which embraces not only the most popular games, such as the N-
person PD, Public Good and Tragedy of the Commons, but also, e.g., the Assurance game,
the Volunteer's Dilemma and their variants with a threshold. In the second part, using the
introduced formalism and the axiomatization, we propose a simple categorization of the
multiperson social dilemmas, based on the formalism of the evolutionary game theory, more
in particular on the number and properties of the equilibria of the corresponding replicator
dynamics (RD) [8,20]. For the considered games with two strategies, the RD is reduced to
one differential equation, with polynomial rhs's. The classification is based on the number
of stable polymorphic equilibria of the RD. In the case of two-player games, it gives only
three standard types of social dilemma games: the PD game, the Chicken Game and the Stag
Hunt game.

In the next section we recall, using a unified notation, various types of multiperson social
dilemmas. In Sect. 3, we formulate the axioms which define social dilemma games, and we
prove that, in particular, all presented multiperson games satisfy the axioms, i.e., they are
social dilemma games. In Sect. 4, we propose the classification of multiperson social dilemma
games. The paper ends with a discussion and suggestions of some directions of research.

2 Multiplayer Social Dilemmas

We consider general N-person strategic games

$$\langle \aleph, \{C,D\}, P_C, P_D \rangle, \tag{1}$$

where $\aleph$ is the set of $N \geq 2$ identical individuals who can choose strategy C: cooperate for
the collective best interest, or D: defect to pursue its own self-interest and receive strategy-
dependent payoffs: $P_C(n)$ for an individual who plays strategy C, and $P_D(n)$ for that who
plays strategy D, where $n \in \{0, 1, ..., N\}$ is the number of players who use strategy C. The
payoff matrix reads

$$
\begin{array}{c|ccccc}
 & \text{C}\cdots\text{C} & \text{C}\cdots\text{CD} & \cdots & \text{CD}\cdots\text{D} & \text{D}\cdots\text{D} \\
\hline
\text{C} & P_C(N) & P_C(N-1) & \cdots & P_C(2) & P_C(1) \\
\text{D} & P_D(N-1) & P_D(N-2) & \cdots & P_D(1) & P_D(0)
\end{array}
\tag{2}
$$

where e.g., $\text{C}\cdots\text{CD}$ stands for the set of $N-2$ C-players and one D-player.

The following examples, commonly known as social dilemma games, are presented in the frame of this unified notation.

I. *N-person Prisoner's Dilemma*

Each of N players chooses one of two strategies: C: cooperate, i.e., to generate a benefit b paying a cost c, where $b > c > 0$, or D: defect, i.e., to do nothing. The benefit b is shared by all other players. The individual payoffs of players who use strategy C, D, are given, respectively, by

$$
P_C(n) = -c + \frac{b(n-1)}{N-1}, \quad n = 1, \dots, N,
$$

$$
P_D(n) = \frac{bn}{N-1}, \quad n = 0, \dots, N-1.
$$

II. *Public Goods*

Each of N players has an amount c of an indivisible goods, and two actions: C: Contribute with c into the common pool (cooperate) or D: do not contribute (defect). After the players have chosen their actions, the amount in the common pool is multiplied by r, $N > r > 1$, and distributed equally among all players. When n players choose C, the final payoff of each cooperator, P_C and each defector, P_D are given, respectively, by

$$
P_C(n) = rnc/N, \quad n = 1, \dots, N,
$$

$$
P_D(n) = rnc/N + c, \quad n = 0, \dots, N-1.
$$

III. *Tragedy of the Commons*

Each of N farmers can keep a cow (strategy D) or not (strategy C) on a common pasture. The utility of a cow is b, and its damage to the environment is c, with $0 < b < c < bN$. All farmers pay evenly the total damage of all the cows kept. Let n be the number of farmers who use strategy C, i.e., who do not keep a cow. Thus, the total damage is $c(N-n)$, and the payoffs from both strategies are:

$$
P_C(n) = -\frac{c(N-n)}{N}, \quad n = 1, \dots, N,
$$

$$
P_D(n) = b - \frac{c(N-n)}{N}, \quad n = 0, \dots, N-1.
$$

For $N = 2$ the games I, II, III are the particular cases of the PD game with (see the beginning of Sect. 4 for the notation) $[R, S, T, P] = [b - c, -c, b, 0]$ for the benefit cost PD, $[r/2 + 1, r, 1, r/2]$ for the Public Good (with normalization $c = 1$) and $[b - c/2, 0, b - c, -c/2]$ for the Tragedy of the Commons.

IV. *Volunteer's Dilemma*

Electricity has gone out in a quarter. Each of N inhabitants can notify the electric company at some cost c (strategy C) or not (strategy D). The benefit of each inhabitant is b if $n \geq 1$ inhabitants notify the company, and 0 otherwise, where $b > c > 0$.

The payoffs read:

$$
P_C(n) = b - c, \quad P_D(n) = b, \text{ if } n \geq 1, \quad P_D(0) = 0.
$$

V. Assurance Game

Each of N players receives a unit as an initial endowment and has two options: to work on a common project (action C) or not to work (action D). To accomplish the project, all players have to work on it, and the final benefit of each of N players is $b > 1$. If at least one player does not work, those who kept working on the project obtain zero. Those who do not work keep 1 as their final payoff. The corresponding payoff functions read

$$P_C(N) = b > 1, \quad P_C(n) = 0 \;\; \text{if } N > n \geq 1, \quad P_D(n) = 1, \quad \text{if } n \geq 0.$$

3 Axiomatization of Multiplayer Social Dilemmas

In this section, we develop an axiomatic derivation of the social dilemma games for general multiplayer games (1) and show that, in particular, the N-person games defined in the previous sections are the social dilemma games.

Definition 1 Multiplayer social dilemma game is the strategic game (1) which satisfies the axioms:

Axiom 1. N C-players are better off than N D-players:

$$P_C(N) > P_D(0), \quad n = 1, \ldots, N - 1. \tag{3}$$

Axiom 2. The payoffs are nondecreasing functions of the number of cooperators:

$$P_C(n) \leq P_C(n + 1), \quad n = 1, \ldots, N - 1. \tag{4}$$

Thus, if a C-player changes to D, the other C-players are not better off then before the change.

$$P_D(n) \leq P_D(n + 1), \quad n = 0, \ldots, N - 2. \tag{5}$$

In other words, if a D-player changes to C, the other D-players are not worse off then before the change.

Axiom 3. Strategy C does not 'weakly' dominate strategy D:

$$\sim \Big(\forall n = 1, \ldots, N \;\; P_C(n) \geq P_D(n - 1) \Big). \tag{6}$$

Otherwise, the players could be tempted to choose C, i.e., there would be no "dilemma." We can rewrite (6) in the form:

$$\exists n \in \{1, \ldots, N\} : P_C(n) < P_D(n - 1).$$

This set of axioms is logically consistent: none of the axiom follows from other two axioms. By straightforward calculations, we check that the multiplayer games I, II, III, IV, V satisfy Axioms 1, 2, 3 for all $N \geq 2$, i.e., they are multiplayer social dilemma games. We note that various other classes of N-person games, popularly described as social dilemmas, also satisfy the axioms. As an example, we consider the N-person PD game with a threshold $M > 0$, defined as follows. Each of cooperators bears a cost $c > 0$ and generates a benefit $b > c$, but only if the number of cooperators $n \geq M$. Otherwise their effort is lost. The benefit b is shared among all other $N - 1$ players. Defectors generate no benefit. Formally, with Θ being a step function: $\Theta(x) = 1$ for $x \geq 0$, $\Theta(x) = 0$ otherwise,

$$P_C(n) = \frac{(n-1)b}{N-1}\Theta(n - M) - c, \quad n = 1, ..., N, \tag{7}$$

$$P_D(n) = \frac{nb}{N-1}\Theta(n - M), \quad n = 0, ..., N - 1. \tag{8}$$

We left to the reader verification of the axioms. Similarly, the axioms are satisfied for the variants of the Volunteer's Dilemma and the Assurance game, in which the payoff from cooperation increases with the number of cooperators, and for analogous games with a threshold.

We introduce a formal definition of greed and fear, characterizing various types of psychological tension in the considered social dilemma games. In the next section, we apply these notions to characterize the social tension in 2-player social dilemmas.

Consider a group of N individuals with $n \geq 2$ C-players.

1. Greed: Greed occurs when a C-player would be better off exploiting other C-players by changing to D: $P_C(n) < P_D(n-1)$.

2. Fear: Fear is present when a C-player X is afraid that if another C-player changes to D, then X receives less, whereas X would be better off by changing to D as well. Formally:

$$\forall n = 2, ..., N: \quad P_C(n) > P_C(n-1) \quad \wedge \quad P_D(n-2) > P_C(n-1). \tag{9}$$

4 Two-Player Social Dilemmas

In this section, we show that the general two-player symmetric games with the payoff matrix

	C	D
C	R	S
D	T	P

for which we use abbreviation $[R, S, T, P]$ admit only three types of social dilemmas. We prove

Lemma 1 *For N = 2 and the general payoff matrix $[R, S, T, P]$, assuming $R \neq T$, $S \neq P$ there exist only three types of social dilemmas: the PD: $T > R > P > S$, the Chicken (Snowdrift) game: $T > R \geq S > P$ and the Stag Hunt game: $R > T \geq P > S$ game.*

We note that with the restriction $R \neq T$, $S \neq P$, in particular the Weak PD game $T > R > P = S$ [18] is not included into the scheme, on the other hand we also avoid other types of nongeneric games, which otherwise would have to be admitted as social dilemma games. The equality $R = S$ allows to classify the two-person Volunteer's Dilemma (see Example IV of Sect. 2) as a particular type of the Chicken game. The Hawk–Dove game $[\frac{v}{2}, 0, v, \frac{v-c}{2}]$, $c > v > 0$ is also a particular case of the Chicken game, with strategy Dove as a cooperative strategy. The weak inequality $T \geq P$ admits as Stag Hunt games, the games in which the payoff of a defector decreases with the number of defectors, as well as the games in which it does not depend on the number of defectors.

Proof For two-player symmetric games, Axioms 1, 2 results in the inequalities:

$$R > P, \quad R \geq S, \quad T \geq P. \tag{10}$$

Axiom 3 can be written in the form: $S < P$ or $R < T$, which can be realized by each of the following three pairs of inequalities:

T1. $S < P \wedge R < T$: together with (10) it gives the PD game.

T2. $S > P \wedge R < T$: together with (10) it gives the Chicken game.

T3. $S < P \wedge R > T$: together with (10) it gives the Stag Hunt game.   □

Using the notions of Greed and Fear, defined in the previous section, we note that for the two-person social dilemma games Greed occurs when $R < T$, whereas Fear occurs when

$R > S$ and $P > S$. Analysis of the payoff matrices of the three social dilemma games shows that in the PD both Greed and Fear are present, in the Chicken game only Greed, and in the Assurance game only Fear are present.

Note that in all these cases not only $R > P$, as required by Axiom 1, but also $T > S$, i.e., in the mixed 2-person group the defector is better off then the cooperator. We can generalize this requirement to general multiplayer games, imposing the additional axiom: In any mixed group (i.e., with at least one D-player and one C-player), D-players are better off than C-players:

$$P_D(n) > P_C(n) \quad n = 1, \ldots, N - 1. \tag{11}$$

We also note that while for $N = 2$ inequality (11) results from Axioms 1, 2, 3, for $N > 2$ it would limit the number of different types of multiplayer social dilemma games.

For the N-player games in which each person plays, with each of $N - 1$ other persons, the general two-player game defined at the beginning of this section, the payoff of a player who plays strategy C or D is defined as, respectively,

$$P_C(n) = \frac{1}{N-1}[R(n-1) + S(N-n)], \quad n = 1, 2, \ldots, N, \tag{12}$$

$$P_D(n) = \frac{1}{N-1}[Tn + P(N-n-1)], \quad n = 0, 1, \ldots, N-1. \tag{13}$$

By straightforward calculations, we prove

Lemma 2 *The N-person games (12), (13), which satisfy $R \geq S$, $T \geq P$, $R > P$, and one of the pair of inequalities T1, T2, T3 in the proof of Lemma 1, satisfy Axioms 1–3.*

5 Classification of Social Dilemma Games

We classify the social dilemma games using the formalism of the evolutionary game theory. We consider an infinite population of individuals who are randomly matched in N-person groups and play at each instant of time a N-person game defined by the payoff matrix (2). The evolution of such system can be described by the replicator equations [8,20]. Let x denote the fraction of cooperators in the population. The replicator equation for the general N-person game reads

$$\frac{dx}{dt} = x(1 - x)W(x), \tag{14}$$

where

$$W(x) = \sum_{k=0}^{N-1} \binom{N-1}{k} x^{N-k-1}(1-x)^k [P_C(N-k) - P_D(N-k-1)]. \tag{15}$$

Our classification of the social dilemma is based on the number and stability properties of the equilibria of (14), i.e., of zeros of the polynomial on the rhs of (14). Let k denote the number of locally attracting polymorphic equilibria of (14), i.e., internal equilibria in the unit interval. We define the following classes of evolutionary social dilemma games.

1. kth order Attractor–Repeller (A–R) dilemma, $k \geq 0$. For $k = 0$, $N = 2$ this is the usual PD, with $x = 0$ globally attracting in $[0, 1)$.
2. kth order Repeller–Repeller (R–R) dilemma, $k \geq 1$. For $k = 1$, $N = 2$ this is the Chicken game, with $x = 0$ and $x = 1$ being repellers.

3. kth order Attractor–Attractor (A–A) dilemma, $k \geq 0$. For $k = 0$ and $N = 2$ this is the Stag Hunt game, $x = 0$ and $x = 1$ being attractors.
4. kth order Repeller–Attractor (R-A) dilemma, $k \geq 1$, with $x = 0$ repeller, $x = 1$ attractor. This class is similar to the class 1 (however, the standard Harmony game, corresponding to $k = 0$ and $N = 2$, is excluded due to Axiom 3).

We classify the general multiperson social dilemmas I–V considered in Sect. 2.

Lemma 3 *The N-person social dilemmas I, II, III are zeroth-order A-R games.*

Proof Straightforward, inserting the relevant payoffs P_C and P_D, using the fact that for a given game the factor $P_C(N - k) - P_D(N - k - 1)$ in (15) is a negative constant, independent of k, and the identity $\sum_{k=0}^{N-1} \binom{N-1}{k} x^{N-k-1}(1 - x)^k = 1$. $\qquad\square$

Lemma 4 *The N-person social dilemma IV is the first-order R–R game. The N-person social dilemma V is the zeroth-order A–A game.*

Proof Straightforward, inserting the relevant payoffs. In the case IV, inserting the payoffs we obtain $W(x) = -c + b(1 - x)^N$. The unique stable polymorphic equilibrium is $x = 1 - [c/b]^{\frac{1}{N-1}}$. In the case V, inserting the payoffs we obtain $W(x) = b - 1 + b(1 - x)^N$. The unique (unstable) polymorphic equilibrium is $x = 1 - [(b - 1)/b]^{\frac{1}{N-1}}$. $\qquad\square$

In this classification we do not, for simplicity, consider situations in which the polynomial $W(x)$ possesses one-dimensional saddles. Below, for brevity, the vector $[P_C(N), P_C(N - 1), \ldots, P_C(1), P_D(N - 1), P_D(N - 2), \ldots, P_D(1), P_D(0)]$ stands for the general N-person game (2). With this notation, for example the three-person game $[2, 2, 0, 3, 1, 1]$ has a unique polymorphic equilibrium, which is a one-dimensional saddle of the corresponding replicator dynamics. Starting from any other initial frequency in the interval $(0, 1)$, the frequency of strategy C decreases. We could call such a game the $\frac{1}{2}$-th order A-R dilemma, due to the following arguments: Increasing $P_D(2)$ in this game from 3, e.g., to 4, we obtain the zeroth-order A-R dilemma, whereas decreasing $P_D(2)$ from 3, e.g., to $2\frac{1}{2}$, we obtain the first-order A-R game, with two polymorphic equilibria: unstable $\frac{2}{7}$ and stable $\frac{6}{7}$. Similarly we could introduce higher-order "fractional" dilemmas, e.g., $1\frac{1}{2}$-order A-R dilemma with a one-dimensional saddle point.

We discuss some other examples of the general classification. The three-person game $[2, 2, 0, 4, 1, 1]$ is a zeroth-order A–R game, with $W(x) = -5x^2 + 4x - 1 < 0$ in (15) and $x = 0$ globally attracting in $[0, 1)$, even though strategy D does not, even weakly, dominate strategy C. Reversing the rows of the payoff matrix of this game, we obtain a zeroth-order R–A, with $x = 1$ globally attracting in $(0, 1]$. The game $[3, 3, 0, 4, 1, 1]$ is a first-order A–R game, with two polymorphic equilibria $\frac{3\mp\sqrt{3}}{6}$, the larger one being an attractor. The game $[2, 1, 1, 3, 2, 0]$ is a first-order R–R game, with the smaller polymorphic equilibrium being an attractor. The game $[3, 0, 0, 2, 1, 1]$ is a zeroth-order A–A game, with a unique, unstable polymorphic equilibrium $\frac{\sqrt{2}}{2}$.

The five-person game $[17, 15, 10, 9, 3, 18, 13, 12, 10, 2]$ is a second-order R–R game, with two stable (and one unstable) polymorphic equilibria. The game $[20, 16, 13, 11, 0, 19, 18, 16, 7, 3]$ is a first-order five-person A–A game, with one stable (and two unstable) polymorphic equilibria.

6 Discussion

In this note, we propose an axiomatic derivation and a simple classification of multiperson social dilemmas played in large populations. Various popular types of social dilemma games, such as the Prisoner's Dilemma, Public Goods, the Tragedy of the Commons, the Volunteer's and the Assurance game, satisfy the proposed axiomatization. Various other social dilemmas, for example the multiperson games with discounted benefit functions, see e.g., [3,7] and references cited therein, can be also included in the considered axiomatic scheme.

In the considered systems, the players are equivalent and they have limited personal capacity. In particular, the players have no memories, they do not make mistakes in their choices of strategies and they can not form collaborative associations. The presented approach does not deal with the problems occurring, e.g., for the two-person repeated games for $T + S > 2R$ when the players can achieve higher average payoffs if they choose C–D and D–C strategy pairs alternately. These problems are left for future research.

The proposed classification is based on the number of stable polymorphic equilibria of the underlining replicator dynamics. The two-player and the higher-order games are treated equally in the proposed classification. However, there are important differences between these two categories of social dilemmas, related to anonymity of decisions, diffusion of the cost of defection and the control of the outcomes of others, cf. e.g., [4,9]. These properties of higher-order social dilemmas should be perhaps taken into account in their more detailed classification. It may also be of interest to study structural changes in the social dilemmas, which may result in transforming the dilemmas, cf. e.g., [17]. One could also study in this respect the role of the price of anarchy, the price of stability and other characteristics of the social dilemmas, cf. e.g., [2] and reference cited therein.

References

1. Akiyama E, Kaneko K (2002) Dynamical systems game theory II: a new approach to the problem of the social dilemma. Phys D Nonlinear Phenom 147(3):221–258
2. Apt KR, Schafer G (2014) Selfishness level of strategic games. J Artif Intell Res 10:207–240
3. Archetti M, Scheuring I (2012) Review: game theory of public goods in one shot social dilemmas without assortment. J Theor Biol 299:9–20
4. Dawes RM (1980) Social dilemmas. Ann Rev Psychol 31:169–193
5. Dawes RM, Messick M (2000) Social dilemmas. Int J Psychol 35(2):111–116
6. Hamburger H (1973) N-person Prisoner's dilemma. J Math Sociol 3(1):27–48
7. Hauert C, Holmes M, Doebeli M (2006) Evolutionary games and population dynamics: maintenance of cooperation in public goods games. Proc R Soc B 273:2565–2571
8. Hofbauer J, Sigmund K (1998) Evolutionary games and population dynamics. University Press, Cambridge
9. Kollock P (1998) Social dilemmas: the anatomy of cooperation. Ann Rev Sociol 24:183–214
10. Komorita SS (1976) A model of the n-person dilemma-type game. J Exp Soc Psychol 12:357–373
11. Lange PV, Balliet DP, Parks CD, van Vugt M (2014) Social dilemmas: understanding human cooperation. Oxford University Press, Oxford
12. Liebrand W (1983) A classification of social dilemma games. Simul Games 14:123–138

13. Liebrand W, Messick D, Wilke H (eds) (2014) Social dilemmas: theoretical issues and research findings. Taylor & Francis, London
14. Macy MW, Flache A (2002) Learning dynamics in social dilemmas. Proc Natl Acad Sci 14 99(10):7229–7236
15. Merlone U, Sandbank DR, Szidarowszky F (2012) Systematic approach to N-person social dilemma games: classification and analysis. Int Game Theory Rev 14(3):1250015
16. Ostrom E (1990) Governing the commons. Cambridge University Press, New York
17. Płatkowski T (2015) Evolutionary coalitional games. Dyn Games Appl. doi:10.1007/s13235-015-0145-3
18. Szabo G, Fath G (2007) Evolutionary games on graphs. Phys Rep 446(4–6):97–216
19. Van Lange PAM, Joireman J, Parks CD, Van Dijk E (2013) The psychology of social dilemmas: a review. Organ Behav Hum Decis Process 120(2):125–141
20. Weibull JW (1995) Evolutionary game theory. MIT Press, Cambridge
21. Wit AP, Wilke AM (1992) The effect of social categorization on cooperation in three types of social dilemmas. J Econ Psychol 13(1):135–151
22. Yao X, Paul JD (1994) An experimental study of n-person iterated prisoner's dilemma games. Informatica 18:435–450