Noname manuscript No.
(will be inserted by the editor)

A Broad Definition of Social Dilemma

Márcia L. Baptista · Pedro A. Santos

Received: date / Accepted: date

Abstract Despite a growing body of research on social dilemmas, a well-accepted formal definition of social dilemma appears to be lacking. This difficulty may stem from the fact that, even though there is a consensus that a social dilemma arises from the collision of individual and collective interest, opinions diverge on the formalization of the notions of individual and collective rationality. We examine the most important definitions of social dilemmas presented to date and propose a novel formal definition. To derive our definition we apply the most well-accepted and strongest notion of collective rationality and the less stringent requirement for the exercise of individual rationality. The proposed definition states that a one-shot non-cooperative game is a social dilemma if there is a rationalizable strategy profile which is Pareto inefficient. By proposing this definition we contend that situations in which there is the possibility but not necessarily the certainty of conflict should be considered social dilemmas. We also show that under this definition, normal-form non-cooperative games which are not social dilemmas in their one-shot versions may pose a dilemma in their repeated versions. A wide range of new games, such as anti-hero and some forms of the iterated deadlock, can now be considered social dilemmas. We also present a taxonomy of 2×2 normal-form non-cooperative social dilemmas.

Keywords Social Dilemma · Individual Rationality · Collective Rationality · Rationalizability · Pareto Efficiency · Taxonomy

1 Introduction

Social dilemmas are at the core of day-to-day social interactions and also of some of the most challenging issues of society related with the environment (global warming, depletion of environmental resources), economics (bank runs, tax evasion) and politics (global war, arms control) to mention a few areas. Despite the importance of the research field, a single unified and well-accepted formal definition of social dilemma has yet to be presented. This may be explained by the fact that, even though it is commonly accepted that a social dilemma arises when there can be a conflict between individual and collective interest, authors often do not formalize their notions of individual and collective rationality or propose conflicting or non-precise formulations of these concepts. It is therefore the aim of this essay to investigate the most relevant and well-known definitions of social dilemmas and to propose a novel and precise definition of social dilemma for normal-form non-cooperative games.

2 From the first notion of social dilemma to current definitions

Gordon Tullock (1974) was the first author to use the term "social dilemma" to describe tension between individual and collective interest. In "The social dilemma: The economics of war and revolution", where Tullock applies economic theory to investigate the dynamics of political conflict, the author ascertains that social dilemmas arise when the rational behaviour of individuals leads to socially inefficient outcomes. A passage of the book is hereafter reproduced to provide a clarification of the concept of social dilemma in Tullock's own words.

> "... conflict uses resources, hence it is socially inefficient, but entering into the conflict may be individually rational for one or both parties. (...) The social dilemma, then, is that we would always

Address(es) of author(s) should be given

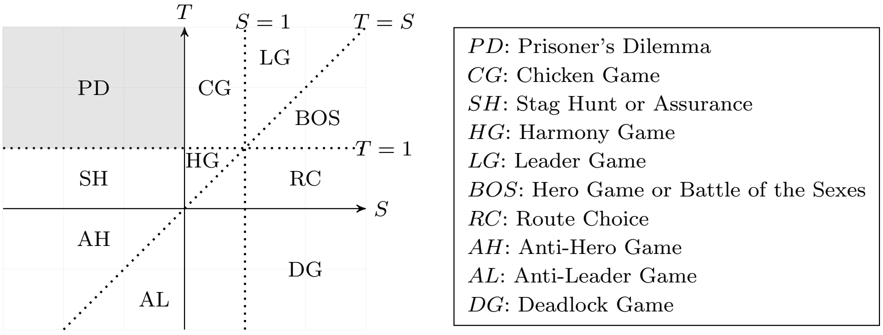


Fig. 1: 2×2 symmetric normal non-cooperative social dilemmas according to Dawes (1975) ($R = 1 > P = 0$ S, T payoff plane)[a b].

[a] Equations 1 and 2 are translated into $T > R \wedge P > S$ and $R > P$ respectively.
[b] Refer to Appendix A.1 for an analysis of the payoff planes.

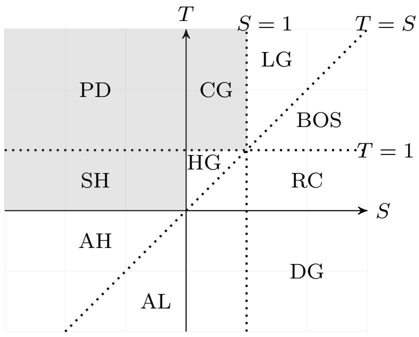


Fig. 2: 2×2 symmetric normal non-cooperative social dilemmas according to Liebrand (1983) ($R = 1 > P = 0$ S, T payoff plane)[a].

[a] In the $R = 1, P = 0$ payoff plane, where Liebrand (1983)'s second property of a Pareto inefficient outcome holds since $R > P$, the first property results in the intersection of two regions: the region where defection is a best response entailing that $T > R \vee S < P$ and the region where defection is a most-threatening strategy implying that $S < R \wedge T > P$.

> be better off collectively if we could avoid playing this kind of negative sum game, but individuals may make gains by forcing such a game on the rest of us." (*The social dilemma: The economics of war and revolution*, p. 2, Tullock, 1974)

Tullock's description of social dilemma was used in the context of a negative sum prisoner's dilemma (PD). However, the author captured the fundamental trait of a social dilemma − the conflict between individual and collective rationality, a general definition on which most authors agree (Bates, 1995; Kollock, 1998; Mak and Rapoport, 2013; Ostrom and Walker, 2005; Sell and Love, 2009; Swedberg, 2001; Van Lange and Liebrand, 1991). Following this preliminary notion, the term social dilemma was first formally defined by Dawes in 1975. Dawes classified a game of 2 strategies and N players as a social dilemma if the following two inequalities would hold simultaneously:

$$D(m) > C(m + 1) \quad (1)$$

$$D(0) < C(N) \quad (2)$$

where $D(x)$ is the payoff of a defector and $C(x)$ is the payoff of a cooperator when x players cooperate.

As it is straightforward to demonstrate, Dawes' definition restricts the set of 2 × N symmetric social dilemmas to prisoner's dilemma games. However, as an analysis of the complete set of 2×2 symmetric games can show (refer to Figure 1), the prisoner's dilemma is only a considerably small subset of games. Accordingly, a great number of strategic games are excluded from the class of social dilemmas according to Dawes' definition. Despite this fact, only in 1983 was a formal definition advanced, enclosing a broader range of games. Liebrand (1983) considered Dawes' definition too stringent in regard to the requirement of a strictly dominant strategy. Hence, the author proposed a less restrictive definition:

> "A social dilemma is defined 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 inefficient outcome."

In Liebrand's definition, the requirement of a strictly dominant strategy (Equation 1) was replaced by the concept of a most-threatening best response strategy (Property 1) (Hamburger, 1973). This definition extended the set of social dilemmas to include stag hunt and chicken games (refer to Figure 2). The admission of these games as social dilemmas is a consideration shared by the vast majority of scholars namely Kollock (1998), Van de Rijt and Macy (2009), Van Lange et al (2013) and Kugler and Bornstein (2013).

		Player2: Stag (C)	Player2: Hare (D)
Player1	Stag (C)	$1, 1$	$-1, \frac{1}{2}$
	Hare (D)	$\frac{1}{2}, -1$	$0, 0$

$R > T > P > S$

Fig. 3: Example of a stag hunt game.

In stag hunt (SH) (refer to Figure 3), in comparison to the prisoner's dilemma, both the strategy of defection and cooperation are individually rational choices. Hence, the dilemma is not as inevitable as in the prisoner's dilemma. According to Liebrand's definition, it is only necessary that the Pareto inefficient outcome may be possible but not certain. This perspective comes into line with Heckathorn's (1996) description of a social dilemma as a "situation in which actions that are individually rational can lead to outcomes that are collectively irrational." The use of the verb "can" denotes the possibility but not the certainty of the inefficient outcome.

Stag hunt also differs from the prisoner's dilemma in the sense that incentives to defect from the cooperative solution are absent. We argue that the existence of an unstable socially desirable outcome adds to the tragedy of the dilemma but is not a fundamental trait of a social dilemma as some authors such as Holt and Roth (2004) contend. In fact, this property is not referred by most authors as essential and games such as stag hunt provide illustrative examples of the possibility of conflict between individual and collective rationality in the absence of such requirement.

		Player2: Swerve (C)	Player2: Dare (D)
Player1	Swerve (C)	$1, 1$	$\frac{1}{2}, 2$
	Dare (D)	$2, \frac{1}{2}$	$0, 0$

$T > R > S > P$

Fig. 4: Example of a chicken game.

The chicken game (CG) differs from the prisoner's dilemma by having an unstable outcome of mutual defection. The stability of the inefficient outcome is referred by some authors such as Kollock (1998) and Swedberg (2001) as a central property of social dilemmas. Nevertheless, we argue that this condition should be perceived as subordinate as Schelling (1973) had already elaborated in the following transcript.

> "So we should probably identify as the generic problem not the inefficient equilibrium of "prisoner's dilemma" or some further reduced subclass, but all the situations in which equilibria achieved by unconcerted or undisciplined action are inefficient − the situations in which everybody could be made better off or the collective total made larger by concerted ... decisions." (Schelling, 1973)

As Schelling notes, the dilemma lies in the possibility of the inefficient outcome which may or may not be a Nash equilibrium. We concur that it is the conflict between individual and social rationality which originates the dilemma and not the inevitability or inability to escape the conflict once reached.

From the previous analysis we argue that in social dilemmas the following conditions should hold true: (1) the inefficient outcome may not be predicted with certainty by the individual rationality principle, (2) the outcomes which dominate the individually rational inefficient outcome may be Nash equilibria and the

(3) inefficient outcome composed of individually rational strategies do not necessarily need to be a Nash equilibrium.

Since Liebrand's definition, a somewhat reduced number of formal definitions of social dilemma have been proposed. Swedberg (2001) advanced the definition of social dilemmas as settings where "strategies, which are individually dominant, converge toward a deficient equilibrium" while Holt and Roth (2004) defined social dilemmas as games where "there is a socially desirable action that is not a Nash equilibrium". Nonetheless, both these definitions contradict the previous conclusions. A more general definition, which is acquiring an undisputed character in the field (Komorita and Parks, 1995; Ostrom and Walker, 2005), is the notion that social dilemmas must be marked by a potential or inevitable conflict between individual and collective rationality. We use this notion as our point of departure to derive a formal definition of social dilemma.

3 Deriving a novel formal definition of social dilemma

If authors seem to agree that a social dilemma arises when individual rationality can be at odds with collective rationality, a precise formulation of the concepts of individual and collective rationality in that context appears to be lacking. We resort to the Pareto principle to define collective rationality in a social dilemma. Even though several notions of social welfare have been proposed (Chang, 2000), the Pareto principle stands as the most normative and authoritative definition of collective rationality. We prefer the strong version of the principle since it represents the strongest possible form of collective rationality allowing us to derive a definition which encloses the largest set of games. Regarding the concept of individual rationality, we apply the notion of rationalizable outcome (Bernheim, 1984; Pearce, 1984). Since rationalizability is the least stringent application of the principle of individual rationality (Bernheim, 1984), by the use of this concept, we are able to extend the current classification of social dilemmas while maintaining a minimal requirement over the collective exercise of individual rationality. The rationalizability of a strategy profile does not imply that players will forcedly choose that course of action however it guarantees the rationality, from an individual perspective, of choosing that course of action. With this formalization of the concepts of individual and collective rationality we advance a definition for a one-shot non-cooperative game.

Definition 1 (One-shot non-cooperative social dilemma) A one-shot non-cooperative game is a social dilemma if there is a rationalizable pure profile which is (strong) Pareto inefficient[1].

The proposed definition imposes restrictions on the nature of the rationalizable profile requiring it to be composed by pure strategies. This follows from the fact that the outcome of a non-trivial mixed profile is infeasible in an one-shot game and hence, its Pareto inefficiency should not be used to classify a one-shot game as a social dilemma.

There may be strategic games with rationalizable profiles which are weak Pareto efficient but are not strong Pareto efficient. Under the proposed definition, such games should be classified as social dilemmas since Pareto improvements can be made once the rationalizable weak Pareto efficient outcomes have been reached. It should also be noted that the definition imposes the rationalizable outcome to not be Pareto efficient, not requiring, however, the existence of a Pareto efficient strategy profile. This particularity is important for games with infinite or not closed strategy sets where they may not exist Pareto efficient profiles without this fact interfering with the inefficiency of the rationalizable outcomes of the game and hence the inherent social dilemma.

		Player2	
		C	D
Player1	C	$1,1$	$8,-1$
	D	$-1,8$	$0,0$

Fig. 5: Example of a deadlock game.

It could be argued that the Pareto dominance of a mixed strategy profile over a rationalizable outcome would be sufficient for a one-shot non-cooperative game to be considered a social dilemma. On the contrary, we contend that the Pareto inefficiency of the rationalizable outcome should only be tested against pure strategy profiles. Our argument stands on the fact that the outcomes of pure strategy profiles are the only ones with a positive probability to occur in one-shot games. Even though mixed strategies may prescribe

[1] We use hereafter the term (strong) Pareto inefficient to refer to outcomes which are not (strong) Pareto efficient.

courses of actions with higher expected payoffs, the application of such strategies in a one-shot game will derive an outcome with a payoff no higher than the payoff outcome of a pure strategy profile. Hence, we claim that mixed strategy profiles should not be used to classify a given rationalizable outcome as strong Pareto inefficient in the context of social dilemmas. To clarify this notion consider the deadlock game depicted in Figure 5. In the game it is possible to apply a mixed strategy with an expected payoff higher than the concrete payoff of mutual cooperation. If both players randomize their strategies, with a probability of $\frac{7}{12}$ of playing cooperation, the expected payoff of each is $\frac{49}{24} \simeq 2$ which is superior to the payoff of 1 obtained from mutual cooperation. Despite the previous result, the rationalizable outcome of $(1, 1)$ of the game should not be considered collectively or socially irrational. Being socially irrational, and in line with the strong Pareto principle, would implicate that there was another outcome in which at least one player could be better off and no other player would be worse off, in comparison with the outcome $(1, 1)$. However, in the one-shot version of the game any mixed strategy profile with an expected payoff superior to $(1, 1)$ will either prescribe the outcome $(-1, 8)$ or $(8, -1)$, none of which Pareto dominates the former. The expected utility outcome of $(\frac{49}{24}, \frac{49}{24})$ is infeasible in a one-shot game since there is only a single play of the game. Since the Pareto criterion should only be applied to outcomes with payoffs with a positive probability to occur, $(1, 1)$ should only be compared with outcomes resulting from the application of pure strategies, namely, $(-1, 8)$, $(8, -1)$ and $(0, 0)$. If there is a Pareto dominant mixed strategy profile with a positive probability to occur, then forcedly there is a Pareto dominant pure strategy profile. Hence, while classifying social dilemmas, the Pareto criterion should be checked against pure strategy profiles in one-shot games.

As referred before, Definition 1 can only be applied to one-shot non-cooperative games. It is however, straightforward to extend the definition to finitely repeated games. Since any finitely repeated game can be represented as a one-shot game, to analyse if a given repeated non-cooperative game is a social dilemma one should represent it as a one-shot game − if the resulting game is a social dilemma then the original game should necessarily be considered a social dilemma. The corollary of the definition follows.

Corollary 1 (Finitely repeated non-cooperative social dilemma) *If a finitely repeated non-cooperative game represented as a one-shot game is a social dilemma then the repeated game is a social dilemma.*

Representing a finitely repeated game as a one-shot game to classify it as social dilemma may prove to be a burdensome task for some games. A more direct and preferable approach is to analyse a more compact representation of the game. This is possible since the strategy profile of a repeated game consisting of the successive play of a rationalizable strategy profile for the one-shot game is also rationalizable given the definition of rationalizable strategy (Pearce, 1984). Since a strategy profile composed of the successive play of a strategy profile which is strong Pareto inefficient in a one-shot game must necessarily be strong Pareto inefficient in a repeated game we can conclude that if a game is a one-shot social dilemma then its finitely or infinitely repeated version is forcedly a social dilemma. Formalization of this corollary follows.

Corollary 2 (Repeated non-cooperative social dilemma) *If a one-shot game is a social dilemma then any finitely (with more than n shots with $n \in \mathbb{N}$) or indefinitely repeated non-cooperative version of the same game is a social dilemma.*

		Player2			
		CC	DD	CD	DC
Player1	CC	$2, 2$	$16, -2$	$9, 0$	$9, 0$
	DD	$-2, 16$	$0, 0$	$-1, 8$	$-1, 8$
	CD	$0, 9$	$8, -1$	$1, 1$	$7, 7$
	DC	$0, 9$	$8, -1$	$7, 7$	$1, 1$

Fig. 6: 2-shot version of game in Figure 5.

From Corollary 1 it follows that certain games which are not social dilemmas as one-shot games may be social dilemmas in their finitely or infinitely repeated versions. An example of such game is the repeated 2×2 non-cooperative deadlock game where $S + T > 2R$. For instance, consider the deadlock game illustrated in Figure 5 − the game is not a social dilemma in its one-shot version but is a social dilemma in its two-shot version as Figure 6 shows. In the two-shot game, the rationalizable profile of collective consistent cooperation yields a worse payoff for all players than other highlighed strategy profiles. The game does not pose an individual or collective dilemma to the participants in the one-shot version since cooperation is the only rational choice from individual and collective perspective in the short-run. However, in the long-run a more rational collective alternative is disregarded: the strategy profile of reciprocity. This type of dilemma can be considered even more acute and difficult to overcome than the traditional prisoner's dilemma since not only the postulate of individual rationality but also the one of collective rationality prescribes a course of action in the short run which does not yield in the long run the best possible outcome for society.

Even though we discussed that mixed strategy profiles should not be considered in one-shot games to classify rationalizable outcomes as inefficient they can provide useful information regarding the existence of a dilemma in the repeated versions of the one-shot game. The following result holds for games with rational payoffs.

Proposition 1 *If a one-shot non-cooperative finite game has a rationalizable profile which is strong Pareto dominated by a non-trivial mixed strategy profile, then the indefinitely repeated version of the game is a social dilemma and there is an $M \in \mathbb{N}$ such that the finitely repeated version of the one-shot game with M is also a social dilemma.*

Proof In this proof we use the notation from Myerson (Section 3.1-3.2 1997). Let

$$\Gamma = (N, (C_i)_{i \in N}, (u_i)_{i \in N})$$

denote the strategic form game, where N is the set of players, C_i is the set of strategies for player i, $C = \bigotimes_{i \in N} C_i$ is the set of profiles and $u_i : C \to \mathbb{R}$ is player's i payoff function. Denote by $\sigma = (\sigma_i)_{i \in N}$ the mixed strategy profile, which we may assume, without loss of generality, to be composed of rational probabilities. We have, for each player i,

$$\sum_{c_i \in C_i} \sigma(c_i) = 1.$$

If s denotes the dominated rationalizable profile, then for any player i

$$u_i(s) < \sum_{c \in C} \left(\prod_{j \in N} \sigma_j(c_j) \right) u_i(c). \qquad (3)$$

There is an M (the least common denominator of the probabilities when written as fractions in the mixed strategy profile) such that

$$M \prod_{j \in N} \sigma_j(c_j) \in \mathbb{Z} \qquad \text{for all } c \in C.$$

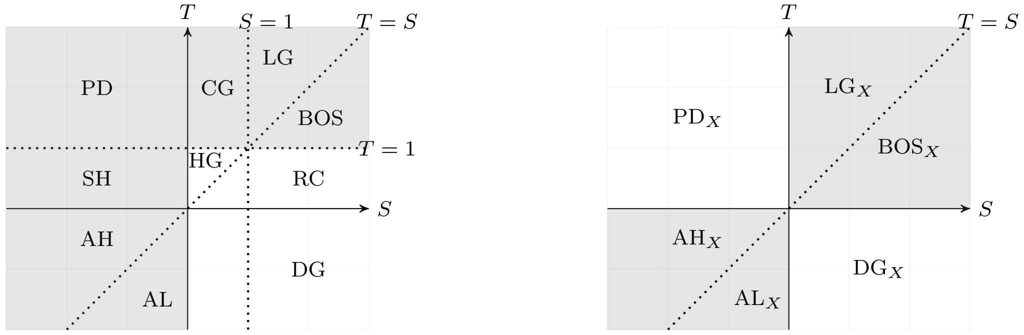
Then there is a repeated game with M shots in which there is a (pure) strategy profile $s*$ for the whole game prescribing at each shot a strategy profile such that in the end the number of times each profile $c \in C$ has been choosen is given by $M \prod_{j \in N} \sigma_j(c_j)$. If one multiplies Equation (3) on both sides by M, we see that the $s*$ profile yields a higher outcome for player i than the rationalizable sucessive pursuit of the one-shot profile s. Accordingly it strong Pareto dominates the former. The repeated game with M rounds is therefore a social dilemma. The indefinitely repeated game based on the one-shot game must also be considered a social dilemma since the player must foresee the worst case scenario which is the possibility that the game will last a multiple of M rounds.

4 Taxonomy of 2 × 2 non-cooperative social dilemmas

With this novel definition a new set of 2 × 2 non-cooperative games, beyond the prisoner's dilemma, stag hunt and chicken game, are considered social dilemmas. Consider Figures 7a and 7b for an exhaustive schematic representation of the set of 2 × 2 one-shot non-cooperative social dilemmas and Figures 8a and 8b for an an identical representation for the set of 2 × 2 repeated non-cooperative social dilemmas.

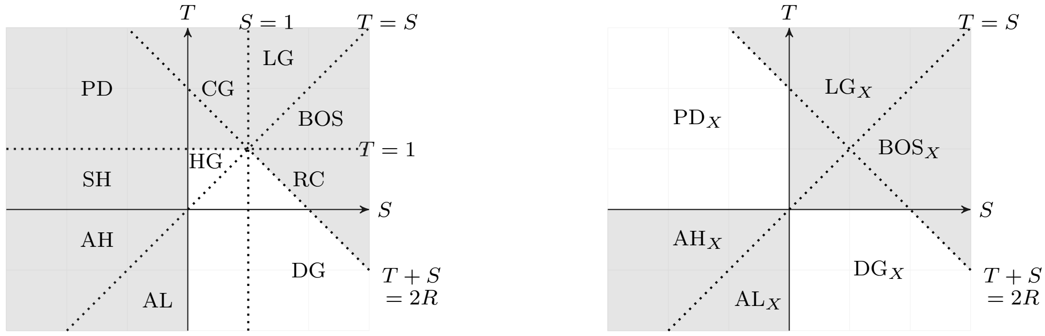
For 2-player games, the set of all rationalizable strategies can be found by iterated elimination of strictly dominated strategies. In the $R = 1, P = 0$ payoff plane (Figure 7a), the set of games where defection is rationalizable is defined by region where $T > R \vee S < P$. In this payoff plane, the collective choice of defection is strong Pareto inefficient. Hence, all games in that region are social dilemmas. In the remaining area of the same plane, defection is strictly dominated and cooperation, when taken collectively, is strong Pareto efficient.

In the $R = P = 0$ payoff plane (Figure 7b), in the region where $T < R \wedge S < R$ cooperation is the only rationalizable strategy and their collective pursuit is strong Pareto efficient. Similarly, in the region where $T < R \wedge S > R$ defection is the only rationalizable strategy and the strategy profile of collective defection is strong Pareto efficient. Games in these two regions are therefore non-social dilemmas. However, in the regions $T < R \wedge S < R$ and $T > R \wedge S > R$ cooperation and defection are both rationalizable and therefore dilemma can arise if players decide to follow cooperation and defection in an anti-coordinated manner.



(a) $R = 1 > P = 0$ S, T payoff plane. (b) $R = P = 0$ S, T payoff plane.

Fig. 7: 2 × 2 symmetric one-shot non-cooperative social dilemmas according to Definition 1.



(a) $R = 1 > P = 0$ S, T payoff plane. (b) $R = P = 0$ S, T payoff plane.

Fig. 8: 2 × 2 symmetric finitely and indefinitely repeated non-cooperative social dilemmas according to Definition 1.

5 Conclusion

Despite the lack of a consensus on the formal definition of social dilemma it is commonly accepted that social dilemmas arise when there can be a conflict between individual and collective rationality. Even though this general definition is well established, authors tend to diverge on the articulation of the concepts of individual and collective rationality. We argue that to derive a formal definition of social dilemmas we should use the less strict well-accepted notion of individual rationality and the most authoritative and indisputable and strongest notion of collective rational reasoning.

Our definition implies in particular, that social interactions in which there is the possibility of individuals pursuing a given course of action leading to conflict, should be considered social dilemmas. Hence, we argue that social dilemmas should be perceived as social settings prone to conflict instead of settings on which conflict is certain. Furthermore, we have shown that if conflict can arise in a set of iterated interactions, then the entire situation must be forcedly considered a social dilemma.

We also show that social dilemmas arise in situations (such as 2×2 iterated deadlock games where $S + T > 2R$) in which players are propelled to choose strategies which are the only possible individual and social rational choices in the short-term with the collective consecutive pursuit of such short-term reasoning deriving however an inefficient outcome in the long-run for all individuals in society.

Our formulation of the concept of social dilemma allows considering a set of games which were not enclosed in previous definitions such as the ones proposed by Dawes (1975) and Liebrand (1983). In the set of 2×2 strategic games, games such as battle of sexes, leader and anti-leader, anti-hero and some forms of the iterated deadlock and route choice, mostly ignored in the theoretical and experimental field of social dilemmas, can now be prone to further investigation.

6 Acknowledgements

This work was partially supported by the INVITE project (ref. UTA-Est/MAI/ 0008/2009) funded by FCT under the UT-Austin/Portugal cooperation agreement and by CEAF-IST, under FCT project PEst-OE/MAT/UI4032/2011.

A Appendix

A.1 Analysis of normalized payoff planes of 2×2 symmetric normal-form non-cooperative games

		Player 2: Strategy 1	Player 2: Strategy 2
Player 1	Strategy 1	1, 1	S, T
	Strategy 2	T, S	0, 0

Fig. 9: Canonical model of the normalized payoff plane where $R > P$.

		Player 2: Strategy 1	Player 2: Strategy 2
Player 1	Strategy 1	0, 0	S, T
	Strategy 2	T, S	0, 0

Fig. 10: Canonical model of the normalized payoff plane where $R = P$.

The whole set of possible 2×2 symmetric games can be analysed solely with the normalized payoff planes where $R > P$ and $R = P$ since the plane where $R < P$ consists of games which are isomorphic to the ones in plane $R > P$.

The first payoff plane (Figure 1) can be derived by considering a game where each player has at his disposal two strategies, strategy 1 and strategy 2, with the supporting payoff matrix being the one depicted in Figure 9 where the individual reward for mutual selection of the first strategy is set to one ($R = 1$) and the individual payoff for mutual selection of the second strategy yields zero ($P = 0$) for both players. The player selecting strategy 1 receives S (suckers payoff) against a player selecting strategy 2 who gets T (temptation to defect). Using this notation, the $R = 1 > P = 0$ S, T payoff plane can be derived. Any 2×2 symmetric game with payoffs R', S', P', T' where $R' > P'$ can be transformed into an incentive structure of the form of the payoff matrix in Figure 9 and thus be projected in the normalized payoff plane by setting $R = 1$, $S = \frac{S'-P'}{R'-P'}$, $T = \frac{T'-P'}{R'-P'}$, $P = 0$.

The same reasoning can be applied to derive the second normalized payoff plane (Figure 7b). The canonical supporting payoff matrix of the plane is depicted in Figure 10 where the individual reward for mutual selection of strategy 1 is set to zero ($R = 0$) and the individual payoff for mutual selection of the second strategy yields zero ($P = 0$) for both players. Any 2×2 symmetric game with payoffs R', S', P', T' where $R' = P'$ can be transformed into an incentive structure of the form of the payoff matrix in Figure 10 and thus be projected in the normalized payoff plane by setting $R = 0$, $S = S' - R'$, $T = T' - R'$, $P = 0$.

It should be noted that the previous transformations do not affect the Pareto efficiency of strategy profiles nor the Nash equilibria of the original games.

To identify and designate the games of the $R > P$ payoff plane we followed the work of Hauert (2002) (prisoner's dilemma, stag hunt, chicken, harmony, leader, battle of the sexes and deadlock games), Tanimoto (2007) (anti-hero and anti-leader) and Stark et al (2008) (route choice). The designation of games in the $R = P$ payoff plane was derived by finding similar games in the $R > P$ plane and assigning the unique subscript X.

References

Bates RH (1995) Social dilemmas and rational individuals. In: Harriss J, Hunter J, Lewis C (eds) The new institutional economics and third world development, Routeledge, pp 27–48

Bernheim B (1984) Rationalizable strategic behavior. Econometrica: Journal of the Econometric Society 52(4):1007–1028

Chang HF (2000) A liberal theory of social welfare: fairness, utility, and the Pareto principle. The Yale Law Journal 110(2):173–235

Dawes RM (1975) Formal models of dilemmas in social decision-making. In: Schwartz MKS (ed) Human judgement and decision processes: Formal and mathematical approaches, New York: Academic Press, pp 87–106

Hamburger H (1973) N-person prisoner's dilemma. Journal of Mathematical Sociology 3(1):27–48

Hauert C (2002) Effects of space in 2×2 games. International Journal of Bifurcation and Chaos 12(07):1531–1548

Heckathorn D (1996) The dynamics and dilemmas of collective action. American Sociological Review 61(2):250–277

Holt C, Roth A (2004) The Nash equilibrium: A perspective. In: Proceedings of the National Academy of Sciences of the United States of America, National Academy of Sciences, vol 101, pp 3999–4002

Kollock P (1998) Social dilemmas: The anatomy of cooperation. Annual Review of Sociology 24:183–214

Komorita SS, Parks CD (1995) Interpersonal relations: Mixed-motive interaction. Annual review of psychology 46(1):183–207

Kugler T, Bornstein G (2013) Social dilemmas between individuals and groups. Organizational Behavior and Human Decision Processes 120(2):191–205

Liebrand W (1983) A classification of social dilemma games. Simulation & Gaming 14(2):123

Mak V, Rapoport A (2013) The price of anarchy in social dilemmas: Traditional research paradigms and new network applications. Organizational Behavior and Human Decision Processes 120(2):142–153

Myerson R (1997) Game Theory: Analysis of Conflict. Harvard University Press

Ostrom E, Walker J (eds) (2005) Trust and reciprocity: Interdisciplinary lessons from experimental research. New York, NY, US: Russell Sage Foundation

Pearce D (1984) Rationalizable strategic behavior and the problem of perfection. Econometrica: Journal of the Econometric Society 52(4):1029–1050

Van de Rijt A, Macy M (2009) The problem of social order: Egoism or autonomy? In: R Thye S, J Lawler E (eds) Altruism and Prosocial Behavior in Groups (Advances in Group Processes, Volume 26), Emerald Group Publishing Limited, pp 25–51

Schelling T (1973) Hockey helmets, concealed weapons, and daylight saving: A study of binary choices with externalities. The Journal of Conflict Resolution 17(3):381–428

Sell J, Love TP (2009) Common fate, crisis, and cooperation in social dilemmas. In: R Thye S, J Lawler E (eds) Altruism and Prosocial Behavior in Groups (Advances in Group Processes, Volume 26), Emerald Group Publishing Limited, pp 53–79

Stark HU, Helbing D, Schönhof M, Hołyst J (2008) Alternating cooperation strategies in a route choice game: Theory, experiments, and effects of a learning scenario. In: Innocenti A, Sbriglia P (eds) Games, Rationality and Behaviour, Houndmills and New York: Palgrave McMillan, pp 256–273

Swedberg R (2001) Sociology and game theory: Contemporary and historical perspectives. Theory and Society 30(3):301–335

Tanimoto J (2007) Dilemma solving by the coevolution of networks and strategy in a 2×2 game. Physical Review E 76(2):021,126

Tullock G (1974) The Social Dilemma: The Economics of War and Revolution. University Publications, Blacksburg, Va.

Van Lange PA, Liebrand WB (1991) Social value orientation and intelligence: A test of the goal prescribes rationality principle. European journal of social psychology 21(4):273–292

Van Lange PA, Joireman J, Parks CD, Van Dijk E (2013) The psychology of social dilemmas: A review. Organizational Behavior and Human Decision Processes 120(2):125–141