Homo Oeconomicus
https://doi.org/10.1007/s41412-022-00124-9

RESEARCH PAPER

Conceptualizing 'Cooperation' and 'Defection' in the Volunteer's Dilemma (and in Social Dilemmas More Generally)*

Werner Raub[1]

Received: 22 December 2021 / Accepted: 12 May 2022

Abstract
We show how 'cooperation' and 'defection' in the Volunteer's Dilemma can be conceptualized in line with established terminology for game-theoretic models of social dilemmas. Commonly employed labels for strategies in the Volunteer's Dilemma are not well in line with our conceptualization. Also, our conceptualization suggests new theory formation and empirical research on the Volunteer's Dilemma.

Keywords Volunteer's Dilemma · Cooperation · Defection · Social dilemma

JEL codes C72 · C92 · H40

* Discussions with Andreas Diekmann, Wojtek Przepiorka, and Thomas Voss on the Volunteer's Dilemma as well as their comments on earlier versions of this paper are gratefully acknowledged. Two reviewers and Hartmut Kliemt, Editor-in-Chief, also offered useful comments and suggestions. These colleagues are not responsible for my arguments, let alone for my mistakes. I benefited from the hospitality and support through a Fernand Braudel Fellowship at the Department of Political and Social Sciences, European University Institute, Florence. No funding was received to assist with the preparation of this article. The author has no competing interests to declare that are relevant to the content of this article.

✉ Werner Raub
w.raub@uu.nl

[1] Department of Sociology/ICS, Utrecht University, Padualaan 14, 3584 CH Utrecht, Netherlands

1 Introduction

Diekmann's (1985, 1986, 1993) Volunteer's Dilemma (VOD) is a seminal and meanwhile prominent example of a game-theoretic model for *N*-person social dilemmas.[1] VOD sheds light on quite some phenomena in social and economic life (see Diekmann 2016: 127–128 for a brief overview and further references). Specifically, VOD allows for instructive analyses of group size effects on individual contributions to collective goods and on collective good production Diekmann, 1985; Raub, 1988: 346–352). Furthermore, research on VOD includes an impressive array of experimental studies, thus linking theoretical modelling and empirical research (for example Diekmann 1986 and Franzen 1995 on group size effects). It is not surprising, therefore, that VOD has become textbook material (Heifetz, 2012: Chap. 13.1; Holt 2019: Chap. 17; Diekmann 2016: Chap. 6) and is sometimes employed as an example in popular science literature (Rosenthal, 2011: 200–203; Pinker 2021: 231–232).

This note addresses a conceptual issue that has been hitherto rather neglected (but see Raub 1988: 346–352). We focus on how to adequately conceptualize 'cooperation' and 'defection' in VOD. Although at first sight merely a terminological question, the issue nevertheless concerns some key research topics on social dilemmas. We first review properties of VOD. The second section summarizes useful and standard conceptualizations of 'cooperation' and 'defection' for game-theoretic models of social dilemmas. We then derive implications for the special case of VOD. The conclusion offers suggestions for further theory formation and empirical research on VOD.

2 Volunteer's dilemma – a brief summary

We consider the standard symmetric VOD (see already Diekmann 1985, 1986 for the following summary; we neglect important variants of VOD, including asymmetry, timing, or incomplete information; see Diekmann 1993; Weesie, 1993, 1994; Weesie & Franzen, 1998; Diekmann & Przepiorka, 2016 on such variants).

VOD is a noncooperative game with *N* actors.[2] They have binary choices. Actors decide simultaneously and independently whether or not to contribute to a collective good. The good is costly and will be provided if and only if at least one actor – a 'volunteer' – contributes. Contributions by two or more actors are feasible. Then each actor pays the full costs of providing the good but contributions of more than one actor do not further improve the utility level of any actor. An actor's costs *K* of contributing to the collective good are *smaller* than the gains *U* from the good. Table 1 summarizes the normal form of VOD. Rows represent each actor's pure strategies, namely, to contribute (CONTR) or not to contribute (DON'T). Columns indicate the number of *other* actors who contribute. Entries in cells are an actor's

[1] For readability and slightly abusing terminology, in the following 'social dilemma' refers to situations in social and economic life that are represented by game-theoretic models *as well as* to those game-theoretic models themselves.

[2] We employ standard game-theoretic terminology and assumptions, as well as some basic results of the theory of noncooperative games (see, for example, Harsanyi 1977 – not a recent but still an authoritative and reliable source, moreover the source for game theory 'background' in Diekmann 1985).

Table 1 The Volunteer's Dilemma ($U>K>0$; $N\geq 2$)

	Number of other actors playing CONTR				
	0	1	2	…	$N-1$
CONTR	$U-K$	$U-K$	$U-K$	…	$U-K$
DON'T	0	U	U	…	U

payoffs (cardinal utility) as a function of his[3] own strategy and the number of other contributing actors.

We denote the strategy combination such that each actor plays the pure strategy CONTR as JOINT CONTR. Analogously, we denote the strategy combination such that each actor plays the pure strategy DON'T as JOINT DON'T. Key properties of VOD that can be seen from Table. 1 or can be derived from the normal form of VOD are the following:

- VOD has no dominant strategy.
- Neither JOINT CONTR nor JOINT DON'T are (Nash) equilibria.
- CONTR is each actor's unique maximin strategy, with JOINT CONTR the unique maximin point.
- VOD has N equilibria in pure strategies. These are the strategy combinations with exactly one volunteer playing the pure strategy CONTR, while all other actors play the pure strategy DON'T. These equilibria involve a bargaining problem, since each actor prefers the equilibria with other actors volunteering to the equilibrium where he himself is the volunteer. Moreover, while VOD is a symmetric game, the N equilibria in pure strategies imply that actors do not play the same strategies.
- VOD has a unique symmetric equilibrium in mixed strategies (see Diekmann 1985: 607 or Raub & Weesie 1992: 17–18 on how to derive this). In this equilibrium, each actor plays CONTR with probability $p^* = 1 - (\frac{K}{U})^{\frac{1}{N-1}}$ (note that $0<p^* < 1$). We denote this equilibrium as JOINT p^*. This is a weak equilibrium (Harsanyi, 1977: 104). Each actor's expected payoff from JOINT p^* is $U-K$ and equal to the maximin payoff so that this equilibrium is also unprofitable (Harsanyi, 1977: 106).[4]

The solution of VOD depends on the specific game-theoretic rationality assumptions one wishes to use. The solution theory developed in Harsanyi (1977: Chaps. 6 and 7) implies that the solution is the maximin point JOINT CONTR (since the game is unprofitable).[5] Any solution theory such that the solution of a noncooperative game must be an equilibrium and also requiring that the solution of a symmetric game is

[3] Throughout, we use 'he' and 'his' to facilitate readability and without intending gender-bias.

[4] For issues concerning assumptions on mixed strategies and equilibria in mixed strategies for noncooperative games see, for example, Osborne & Rubinstein (1994: passim) and, specifically for VOD, Tutić (2014).

[5] Roughly, Harsanyi's solution theory yields an equilibrium (or a set of equilibria) as the solution, if certain stability criteria are fulfilled (for example, the game is 'profitable' in the sense that the solution yields more than his maximin payoff for each actor), while otherwise the solution is such that the actors employ maximin strategies.

symmetric would imply that JOINT p^* is the solution.[6] Note that the payoff vector associated with the solution is the same in both cases.[7]

Concerning terminology used for VOD in much earlier literature, note that CONTR has been routinely labeled ‘cooperation’, while DON’T has been routinely labeled ‘defection’ (see already Diekmann 1985: 606, 1986: 188, 2016: 125 and more recently Diekmann & Przepiorka 2016: 1315). These labels suggest that common terminology for game-theoretic models of social dilemmas is employed. We now consider in more detail how to conceptualize ‘cooperation’ and ‘defection’ for game-theoretic models of social dilemmas in general as well as for the specific case of VOD.

3 Social dilemmas, cooperation, and defection

While conceptualizations of ‘social dilemma’ slightly differ in some respects (see, for example, Harsanyi 1977: 276–280; Dawes 1980; Kollock, 1998; Raub et al., 2015), it is quite standard to require that a social dilemma is anyway a game with a Pareto-suboptimal solution. This means that there is a feasible strategy combination for such a game that is Pareto-superior to the solution in the sense that it is associated with higher payoffs for at least some actors than their solution payoffs, while not being associated with lower payoffs for any actor. Diekmann (1985) likewise refers to such a conceptualization.

Concerning ‘defection’ and ‘cooperation’ in a social dilemma, it is useful to first characterize strategy *combinations* that represent defection and, respectively, cooperation by *all* actors. *Individual* defection and *individual* cooperation are then an actor’s individual strategies according to these strategy combinations. We will see below that and why it makes sense to develop terminology in this way, rather than trying to define individual defection and individual cooperation without reference to strategy combinations and, hence, without reference to *other* actors’ strategies.[8]

A strategy combination that is a solution of a social dilemma – and is thus Pareto-suboptimal – is then a strategy combination that represents *defection* by all actors. An obvious requirement for a strategy combination that represents *cooperation* by all actors is then that such a strategy combination is Pareto-superior to the solution. It is useful to require furthermore that a strategy combination that represents cooperation by all actors is not only Pareto-superior to the solution but is also Pareto-

[6] This is a good example for the ‘theory dependence’ of claims about games in general and more specifically about game-theoretic models of social dilemmas (see Raub et al., 2015: 605–608).

[7] We briefly mentioned in the introduction that VOD is often employed for analyses on group size effects. This work typically assumes JOINT p^* as the solution. One can then derive the predictions that the individual probability of choosing CONTR and the probability of collective good production (that is, at least one actor chooses CONTR) decline with increasing group size N. Results of experiments often support that increasing group size decreases the individual probability of choosing CONTR, while not supporting that increasing group size decreases the probability of collective good production (see, for example, Diekmann 1986, Franzen, 1995, and Tutić 2014).

[8] We stress that we focus on how ‘defection’ and ‘cooperation’ are typically conceptualized *in the literature on social dilemmas* and what these conceptualizations imply for the special case of VOD. For a much broader discussion of the concept of ‘cooperation’, see for example Brennan & Sayre-McCord (2018).

optimal. Then, there is no other feasible strategy combination that is Pareto-superior to a strategy combination that represents cooperation by all actors. Requiring Pareto-optimality is consistent with common terminology for social dilemmas (see Raub et al., 2015) and ensures that a strategy combination that represents cooperation by all actors is not itself Pareto-suboptimal. Furthermore, requiring Pareto-superiority together with Pareto-optimality is well in line with Rapoport's (1974) intuitive characterization of a social dilemma. Rapoport focuses on the tension between 'individual rationality' in the sense of behavior according to rationality postulates underlying the game's solution and 'collective rationality' in the sense of Pareto-optimality. Coleman (1994: 168) distinguishes in a similar vein between 'social equilibrium', namely, a Nash equilibrium, and 'social optimum', namely, a Pareto optimum, stressing that these are distinct concepts and that a social equilibrium need not be Pareto-optimal. It seems straightforward to build terminology so that cooperation by all actors satisfies collective rationality à la Rapoport and, respectively, is a social optimum à la Coleman. Note that Diekmann & Przepiorka (2016: 1311), in line with Rapoport and Coleman, characterize collective rationality as follows: 'Collective rationality means that actors, had they an opportunity to communicate and agree on a binding contract, should agree on a combination of actions leading to a welfare-enhancing outcome.' It follows from an elementary rationality assumption for such agreements, namely, 'joint efficiency' (for example, Harsanyi 1977: 198), that Pareto-optimality is one of the properties of agreements. Often, but not always, a cooperative strategy combination will not be an equilibrium.

Much theoretical and empirical research on social dilemmas is on conditions and mechanisms that foster cooperation and mitigate defection. This indicates that conceptualizing 'cooperation' and 'defection', while a terminological issue to begin with, is closely connected with problems of theory formation and empirical research.

3.1 Cooperation and defection: strategy combinations and individual strategies

It is useful to note that our conceptualization implies that, in general, it can depend on the strategies of other actors if a given strategy of an actor represents individual defection. Likewise, in general, it can depend on the strategies of other actors if a given strategy of an actor represents individual cooperation. A pure coordination game like in Table. 2 is an example that makes this clear. The example likewise highlights that and why one should first characterize strategy *combinations* that represent defection and, respectively, cooperation by *all* actors, then defining individual defection and individual cooperation as an actor's individual strategies according to these strategy combinations.

In the game in Table. 2, actors have completely identical interests. The game has two equilibria in pure strategies, (TOP, LEFT) and (DOWN, RIGHT). Both equilibria are Pareto-optimal, both actors are indifferent between these equilibria, and both equilibria are associated for each of the actors with the highest payoff that is feasible at all in this game. However, this leaves the question open how actors – without communication and without repeated play – can coordinate their choices. There is also a mixed equilibrium such that each actor plays each of his two pure strategies with

Table 2 Pure Coordination

		Actor 2	
		LEFT	RIGHT
Actor 1	TOP	1, 1	0, 0
	DOWN	0, 0	1, 1

probability ½. This equilibrium is Pareto-suboptimal and less attractive for each actor than the two equilibria in pure strategies.

A solution theory such as Harsanyi's (1977) selects the equilibrium in mixed strategies as the solution, since it satisfies – other than the two pure strategy equilibria – certain stability criteria. Under such a solution theory, the game *is* a social dilemma. Then, the mixed strategy equilibrium would represent defection by all actors, while (TOP, LEFT) as well as (DOWN, RIGHT) would represent cooperation by all actors. Hence, whether an actor's pure strategy represents individual cooperation depends on the strategy of the other actor. Of course, there are also special cases without such complications. For example, in the Prisoner's Dilemma each actor has a dominant strategy, the combination of these dominant strategies is the Pareto-suboptimal solution of the game, and there is a unique strategy combination that is Pareto-superior to the solution as well as Pareto-optimal. In this case, one can say that each actor has a unique strategy that represents individual defection as well as a unique strategy that represents individual cooperation. In this sense, one can 'neglect' for the Prisoner's Dilemma that individual defection and cooperation should be defined *relative to* the other actor's strategy.[9] Furthermore, note that while cooperative strategy combinations are not an equilibrium in many social dilemmas such as the Prisoner's Dilemma, our example of a pure coordination game also shows that cooperative strategy combinations can be equilibria in some social dilemmas.

3.2 'Third party effects' of cooperation and defection

It is also useful to briefly consider issues related to 'third party effects' of cooperation and defection in social dilemmas. First, in the special case of the Prisoner's Dilemma, it is tempting to interpret an actor's defection as opportunistic behavior in the sense of being harmful and exploiting the other actor. However, such an interpretation need not be plausible for all social situations for which the Prisoner's Dilemma is a reasonable formal model. Namely, an assessment of actors' behavior in a Prisoner's Dilemma could and possibly should also account for societal effects of their behavior. For example, from a societal perspective, cooperation of the two criminals in the original story underlying the Prisoner's Dilemma, as well as other instances of cooperation between criminals or cooperation between members of a cartel, can be considered to have detrimental effects for third parties not involved themselves as actors in the game. Therefore, in such situations, defection, rather than cooperation, might be considered desirable from a societal perspective. Conversely, provided that external effects for third parties can be neglected or when cooperation has even posi-

[9] It would be interesting, but is beyond the scope of this paper, to study necessary and sufficient conditions such that each actor has a unique strategy that represents individual defection as well as a unique strategy that represents individual cooperation in a social dilemma.

tive external effects for third parties, it may be plausible to interpret an actor's defection in the Prisoner's Dilemma as opportunistic.

Second, it seems useful to note that related assessments for other social dilemma games can anyway differ from the case of the Prisoner's Dilemma. The pure coordination game is again an example. Under Harsanyi's (1977) solution theory, the mixed strategy equilibrium is the solution of the pure coordination game and represents defection of both actors. However, even when abstracting from possible third party-effects, one would not easily interpret playing the mixed equilibrium strategy as a case of opportunistic behavior. The Trust Game (for example, Dasgupta 1988) and the Investment Game (Berg et al., 1995) are further examples. In these games, plausible solution theories imply for the Trust Game that no trust is placed by the trustor, while the trustee would abuse trust. For the Investment Game, the solution is such that the trustor sends nothing and the trustee would not return anything. These equilibria are Pareto-suboptimal and represent defection by all actors – both games are examples of social dilemmas. In both games, again assuming away third party-effects, the equilibrium strategy of the trustee can be considered a rather clear case of opportunistic behavior, exploiting the trustor's trustfulness. However, the equilibrium strategy of the trustor implies protection against opportunism of the trustee, rather than an attempt to increase the trustor's payoff through exploitation of the trustee. Hence, these games highlight that in social dilemmas a strategy combination representing defection by all actors can be a strategy combination such that one actor's strategy does represent opportunism, while the other actor's strategy 'defends' against opportunism.

While the Prisoner's Dilemma is one specific paradigmatic example of a social dilemma, our discussion also indicates that *specific* properties of defection and cooperation in the special case of the Prisoner's Dilemma need not be *generic* properties of defection and cooperation in social dilemmas. More generally, one should observe that the key point of a game-theoretic perspective on social dilemmas is that individually rational behavior, due to strategic interdependence, can imply collective irrationality in the sense of Pareto-suboptimality (Rapoport) or that a social equilibrium need not be a social optimum (Coleman). While the established labels 'defection' and 'cooperation' may have normative connotations, it is important that the key point of 'unintended consequences of goal-directed behavior' is carefully distinguished from assessments of individual behavior in terms of opportunism and from assessments of the macro-level consequences of individual behavior in terms of effects for third parties not themselves involved as actors in the dilemma.

4 Implications for VOD

Is VOD a social dilemma in the sense that the solution of VOD is Pareto-suboptimal? Note, first, that whether or not VOD is a social dilemma does not depend on whether we assume JOINT CONTR or JOINT p^* as the solution of VOD. After all, the payoff vector associated with JOINT CONTR is the same as the payoff vector (in terms of expected payoffs) associated with JOINT p^*.

Joint randomization (Harsanyi, 1977: 97) would allow for determining one volunteer by a chance mechanism, while all other actors would use DON'T. This would yield expected payoff $U - K/N > U - K$ for each actor.[10] However, since VOD is a noncooperative game, jointly randomized strategies are not feasible (see already Diekmann 1985: 607).[11]

Each of the N asymmetric equilibria in pure strategies is associated with payoffs $U - K$ for the volunteer and payoffs U for all other actors. Therefore, each of these equilibria is weakly Pareto-superior to the solution: the volunteer is not worse off than in the solution, while all other actors are better off. Each of these equilibria is also Pareto-optimal.

There are no feasible strategy combinations in VOD with (expected) payoffs $U - K/N$ for each actor. Consider now the symmetric strategy combination such that each actor plays CONTR with probability $p^{**} = 1 - (\frac{K}{NU})^{\frac{1}{N-1}}$ (note that $p^{**} > p^{*}$). We denote this strategy combination as JOINT p^{**}. Expected payoffs associated with JOINT p^{**} are $U - qK$ for each actor, with $K/N < q < 1$ and therefore also $U - K/N > U - qK > U - K$.[12] Hence, JOINT p^{**} is Pareto-superior to the solution of VOD. Diekmann (1985: 608) pointed out that JOINT p^{**} is also a Pareto-optimal feasible strategy combination for VOD. Referring to Hofstadter (1983), Diekmann (1985: 608) suggested the label 'superrational' for JOINT p^{**}, since actors maximize their expected payoffs under the restriction of Kant's categorical imperative.[13] JOINT p^{**} is not an equilibrium of VOD: DON'T is each actor's best-reply strategy if all other actors play CONTR with p^{**} and it is useful to note that $U - K/N$ is an actor's expected payoff when playing DON'T, while all other actors play CONTR with p^{**} (Raub, 1988: 350). Such an actor's payoff is thus the same as in the case of joint randomization as described above.

Using these results, there is a straightforward *conceptualization of 'defection by all actors' for VOD*. Namely, depending on the solution theory employed, 'defection by all actors' should be conceptualized as the maximin point JOINT CONTR or, respectively, as the mixed symmetric equilibrium JOINT p^{*}. In any case, it would be misleading to consider JOINT CONTR as representing collective rationality or as representing 'cooperation by all actors' in VOD. Also, while JOINT DON'T is of course Pareto-suboptimal and also Pareto-inferior to the maximin point as well as to the mixed symmetric equilibrium, JOINT DON'T is not a maximin point and not an equilibrium. JOINT DON'T is therefore not a solution of VOD and it would be misleading to consider JOINT DON'T as conceptualizing 'defection by all actors' in VOD.

The asymmetric equilibria in pure strategies as well as JOINT p^{**} are Pareto-superior to the solution of VOD and are Pareto-optimal. However, as has been pointed out,

---

[10] With costless side payments, each actor's payoff could be $U - K/N$ not only in terms of the actor's expectation.

[11] The observation that jointly randomized strategies are not feasible in noncooperative games also shows that the characterization of collective rationality in Diekmann & Przepiorka (2016: 1311) that we already mentioned above should be understood to refer to agreements on strategy combinations that are *feasible* in a noncooperative game.

[12] See Diekmann (1985: 608) for the explicit expression for q.

[13] We abstain from discussing the pros and cons of the label 'superrational'.

the asymmetric equilibria in pure strategies are associated with bargaining problems. According to the well-known symmetry postulate (Harsanyi, 1977: 198), rational bargaining yields equal payoffs to all actors in symmetric games. Therefore, standard terminology for social dilemmas and the characterization of collective rationality in Diekmann & Przepiorka (2016: 1311) strongly suggest JOINT p^{**} as the *conceptualization of 'cooperation by all actors' for VOD*.

Note that it might seem tempting to consider a strategy as 'cooperative' if it increases the value of a collective good under at least some conditions and never reduces the value of the collective good.[14] However, a conceptualization along this line seems somewhat ad hoc in the sense of being tailor-made for VOD. Furthermore, it would yield counter-intuitive results for other social dilemmas such as pure coordination games like in Table. 2. In a pure coordination game, the collective good would be 'successful coordination' in the sense of either (TOP, LEFT) or (DOWN, RIGHT). However, whether an actor's strategy is cooperative in the sense of contributing to successful coordination depends on the strategy of the other actor and thus there is no strategy in a pure coordination game that increases the value of a collective good under at least some conditions and never reduces the value of the collective good (similar problems would occur with respect to other social dilemma games as well, such as bargaining dilemmas or games with indifference problems; see, for example, Harsanyi 1977: 278–280).

Summarizing, we have seen that in social dilemma games it can depend on the strategy of other actors if a given strategy of an actor represents individual defection. This is also the case for individual cooperation. We have furthermore seen that and why careful scrutiny is appropriate when considering cooperation and defection as possible instances of opportunism. What about VOD in these respects? First, it is theory dependent whether we assume JOINT CONTR or JOINT p^* as the solution of VOD. Namely, it depends on the solution theory applied whether JOINT CONTR or JOINT p^* is the solution of VOD. However, *given* the solution theory, individual defection is uniquely defined. Also, and even irrespective of the two solution theories we considered, individual cooperation is also uniquely defined for VOD. Second, concerning opportunism, once again assuming away complications due to external effects for actors not themselves involved in the game, it seems implausible to maintain that CONTR, while an instance of individual defection under Harsanyi's (1977) solution theory, represents opportunistic behavior. Again, we see that the tension between individual and collective rationality or social equilibrium and social optimum needs to be distinguished from assessments of individual behavior in terms of opportunism and of the macro-level consequences of individual behavior.

5 Conclusion, including implications for theory formation and empirical research on VOD

Employing established terminology for social dilemmas, we have argued that 'defection' in VOD can best be understood as either all actors playing the pure strategy

[14] I owe this idea to Andreas Diekmann (email-communication, May 11, 2019).

CONTR or all playing the symmetric mixed equilibrium strategy p^*. 'Cooperation' in VOD can best be understood as all actors playing the 'superrational' strategy p^{**}. This suggests to avoid the labels 'cooperation' and 'defection' for the pure strategies CONTR and DON'T. In various respects, our discussion has likewise indicated that notions on the Prisoner's Dilemma do not always generalize straightforwardly to other dilemma games. To repeat, these are terminological issues and such issues should not get too much attention. Still, our discussion concerns not only purely verbal issues merely on labels but also concerns bringing terminology for VOD in line with established terminology for social dilemmas.

While we considered exclusively the standard symmetric VOD, it goes without saying that a systematic comparison with other versions of VOD, including asymmetry, timing, incomplete information, and the like, might be worthwhile. Also, it would be useful to compare the VOD-properties on which we focused to those of games modeling the production of other step-level collective goods.

Discussions of conceptual issues should have ramifications for theory formation and empirical research. In our case, there are implications for theory formation and empirical research on the *repeated* VOD, including the emergence of norms of cooperation in the repeated VOD. Diekmann & Przepiorka (2016) is a seminal theoretical and empirical contribution exploring 'turn-taking' in the repeated VOD and the emergence of norms of 'taking turns': playing CONTR one after the other in subsequent rounds. Of course, while turn-taking would be beneficial for the actors in the repeated VOD compared to the solution of the one-shot VOD, turn-taking comes with bargaining problems for the actors, at least under the assumption of discounting: actors would still prefer other actors to start with playing CONTR in early rounds and to play CONTR themselves only in later rounds. Our conceptualization of cooperation in the VOD suggests further questions on the repeated VOD. Repeated play of the 'superrational' strategy p^{**} by all actors would avoid bargaining problems but usually actors cannot directly observe if other actors play mixed strategies. They can observe the *behavior* of other actors, not the strategy 'generating' the behavior. Hence, standard 'conditional cooperation' in repeated games becomes problematic for the repeated VOD: when an actor's deviation from employing p^{**} is not directly observable, how to credibly threaten deviations? Fudenberg & Maskin (1986) includes a discussion of these issues, showing when and how such problems could be solved in principle. It would be interesting to explore testable implications of a theoretical approach along these lines and to explore if such implications differ from hypotheses on turn-taking as in Diekmann & Przepiorka (2016). Also, truthful information about playing mixed strategies could be made feasible in a suitable experimental design, using a variant of Selten's (1967) strategy method (see also Roth 1995: 320–323). Then, one could empirically investigate if actors' behavior in a repeated VOD differs depending on what they can observe and know with respect to other actors' behavior and 'underlying' strategies.

References

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*, 122–142

Brennan, G., & Sayre-McCord, G. (2018). On 'cooperation'. *Analyse & Kritik*, *40*, 107–130

Coleman, J. S. (1994). A rational choice perspective on economic sociology. In N. J. Smelser, & R. Swedberg (Eds.), *The Handbook of Economic Sociology* (pp. 166–180). Princeton, NJ: Princeton University Press

Dasgupta, P. (1988). Trust as a commodity. In D. Gambetta (Ed.), *Trust: Making and Breaking Cooperative Relations* (pp. 49–72). Oxford: Blackwell

Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology*, *31*, 169–193

Diekmann, A. (1985). Volunteer's Dilemma. *Journal of Conflict Resolution*, *29*, 605–610

Diekmann, A. (1986). Volunteer's Dilemma. A social trap without a dominant strategy and some empirical results. In A. Diekmann, & P. Mitter (Eds.), *Paradoxical Effects of Social Behavior* (pp. 187–197). Heidelberg: Physica

Diekmann, A. (1993). Cooperation in an asymmetric Volunteer's Dilemma game. *International Journal of Game Theory*, *22*, 75–85

Diekmann, A. (2016). *Spieltheorie. Einführung, Beispiele, Experimente* (4th ed.). Reinbek bei Hamburg: Rowohlt

Diekmann, A., & Przepiorka, W. (2016). 'Take one for the team!' Individual heterogeneity and the emergence of latent norms in a Volunteer's Dilemma. *Social Forces*, *94*, 1309–1333

Franzen, A. (1995). Group size and one shot collective action. *Rationality and Society*, *7*, 183–200

Fudenberg, D., & Maskin, E. (1986). The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, *54*, 533–554

Harsanyi, J. C. (1977). *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge: Cambridge University Press

Heifetz, A. (2012). *Game Theory: Interactive Strategies in Economics and Management*. Cambridge: Cambridge University Press

Hofstadter, D. R. (1983). The calculus of cooperation is tested through a lottery. *Scientific American*, *6*, 14–18

Holt, C. A. (2019). *Markets, Games, and Strategic Behavior. An Introduction to Experimental Economics* (2nd ed.). Princeton, NJ: Princeton University Press

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

Osborne, M. J., & Rubinstein, A. (1994). *A Course in Game Theory*. Cambridge, MA: MIT Press

Pinker, S. (2021). *Rationality. What It Is. Why It Seems Scarce. Why It Matters*. Allen Lane

Rapoport, A. (1974). Prisoner's Dilemma – recollections and observations. In A. Rapoport (Ed.), *Game Theory as a Theory of Conflict Resolution* (pp. 18–34). Dordrecht: Reidel

Raub, W. (1988). Problematic social situations and the 'large-number dilemma'. *Journal of Mathematical Sociology*, *13*, 311–357

Raub, W., Buskens, V., & Corten, R. (2015). Social dilemmas and cooperation. In N. Braun, & N. J. Saam (Eds.), *Handbuch Modellbildung und Simulation in den Sozialwissenschaften* (pp. 597–626). Wiesbaden: Springer VS

Raub, W., & Weesie, J. (1992). *Sociological applications of game theory. Mimeo*. Utrecht University

Rosenthal, E. C. (2011). *The Complete Idiot's Guide to Game Theory*. New York: Penguin

Roth, A. E. (1995). Bargaining experiments. In J. H. Kagel, & A. E. Roth (Eds.), *The Handbook of Experimental Economics* (pp. 253–348). Princeton, NJ: Princeton University Press

Selten, R. (1967). Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments. In H. Sauermann (Ed.), *Beiträge zur experimentellen Wirtschaftsforschung (Band 1)* (pp. 136–168). Tübingen: Mohr

Tutić, A. (2014). Procedurally rational volunteers. *Journal of Mathematical Sociology*, *38*, 219–232

Weesie, J. (1993). Asymmetry and timing in the Volunteer's Dilemma. *Journal of Conflict Resolution*, *37*, 569–590

Weesie, J. (1994). Incomplete information and timing in the Volunteer's Dilemma. A comparison of four models. *Journal of Conflict Resolution*, *38*, 557–585

Weesie, J., & Franzen, A. (1998). Cost sharing in a Volunteer's Dilemma. *Journal of Conflict Resolution*, *42*, 600–618