21 Social Dilemmas and Cooperation*

Werner Raub[a], Vincent Buskens[a,b] and Rense Corten[a,c]

[a] Utrecht University

[b] Erasmus University Rotterdam

[c] Tilburg University

Summary. Social dilemmas (sometimes referred to as "problems of collective action", "tragedy of the commons", or "public goods problems") are situations with strategically interdependent actors such that individually rational behavior leads to an outcome that is less desirable for each actor than had they cooperated. In this chapter, we provide an overview of models of social dilemmas and cooperation in social dilemmas that use game-theoretic tools. We first review examples of social dilemmas and formal modeling of such dilemmas. We distinguish between dilemmas that involve two actors and those that can involve more than two actors. We also discuss why the conceptualization of "social dilemma" is "theory dependent". Second, we review mechanisms that can induce cooperation in social dilemmas. Cooperation of rational actors in a social dilemma requires that the dilemma is "embedded" in a more complex game. We discuss models for different types of embeddedness. The chapter likewise includes a sketch of models of cooperation based on social preferences and of simulation studies as an alternative to game-theoretic analysis.

1 Introduction

Actors *cooperate* when they behave so that the outcome is better for each of them than the situation where all or at least some actors abstain from cooperation. A "*social dilemma*" is a situation with strategically interdependent actors such that at least some actors face individual opportunities and incentives to abstain from cooperation (to "*defect*"), while compared to the cooperative outcome all actors are worse off when

* Comments by an anonymous reviewer are gratefully acknowledged. Support for Raub was provided by NWO (PIONIER-program "The Management of Matches"; grants S 96-168 and PGS 50-370) and for Buskens by Utrecht University (High Potential-program "Dynamics of Cooperation, Networks, and Institutions"). Raub acknowledges the hospitality of Nuffield College, University of Oxford. For some parts of this chapter, we use materials from Buskens & Raub (2013).

S. 597–626 in: Norman Braun & Nicole J. Saam, Hg. (2014). Handbuch Modellbildung und Simulation in den Sozialwissenschaften. Wiesbaden: Springer VS

actors follow those incentives. More technically, when actors defect, the outcome is Pareto-suboptimal, while cooperation is Pareto-optimal and a Pareto-improvement compared to the outcome when actors defect. Using game theory (see Tutić's chapter in this Handbook and a textbook such as Rasmusen 2007 for game-theoretic terminology and assumptions) to make the notion precise, a social dilemma game is a non-cooperative game with a "solution" – the strategy combination played by rational actors (this will usually be a Nash equilibrium that satisfies additional criteria such as subgame perfection) – that is Pareto-suboptimal. Cooperation indicates a strategy combination that is associated with an outcome of the game that is Pareto-optimal and a Pareto-improvement compared to the solution, while cooperation is typically not an equilibrium outcome so that there are indeed actors with an incentive to deviate.[1] Thus, in Rapoport's (1974) intuitive characterization, there is a tension between *individual rationality* (see Saam & Gautschi's chapter in this Handbook) in the sense of incentive-guided and goal-directed behavior on the one hand and *collective rationality* in the sense of Pareto-optimality on the other. "Social dilemma" is a label commonly used in social psychology and sociology. In other disciplines such as political science and economics, social dilemmas are often referred to as "problems of collective action", "tragedy of the commons" or "public goods problems" (Ledyard 1995: 122).

Social dilemmas and cooperation are revered topics in social science theory. First, they are a paradigmatic example of unintended consequences of goal-directed behavior: actors try to further their own interests and in doing so produce an outcome that is worse for all than an outcome they could have obtained by cooperating. Also, Pareto-optimality and, respectively, suboptimality are paradigmatic examples of macro-consequences of behavior in interdependent settings in the sense that cooperation and defection are phenomena on the micro-level of individual behavior, while Pareto-optimality and -suboptimality are properties of the social system formed by the actors. Hence, theories and models of social dilemmas and cooperation exemplify the study of micro-macro links in social science (see Raub et al. 2011).

Second, social dilemmas are closely related to Hobbes' ([1651] 1991: chap. 13) "naturall condition of mankind" with interdependent actors in a world of scarcity without binding and externally enforced contracts so that actors can end up in the "warre of every man against every man", while in a peaceful situation – in our terminology: cooperation – everybody would be better off. Parsons' (1937) considered the Hobbesian "problem of order" as "the most fundamental empirical difficulty of utilitarian thought" (1937: 91) and posed the challenge to specify conditions such that rational, i.e., incentive-guided and goal-directed actors can cooperate and thus avoid to end up in the situation vividly described by Hobbes as life being "solitary, poore, nasty, brutish, and short". Durkheim ([1893] 1973, book I: chap. 7), in his analysis of the division of labor in society, likewise discussed a social dilemma, albeit using other terminology. His topic comprises the limits of the contractual governance of economic transactions. Governing a transaction exclusively via a contract would require

[1] This conceptualization has been suggested by Raub & Voss (1986), derives from Harsanyi's (1977) meanwhile classic treatment, and is closely related to various other approaches (see Van de Rijt & Macy 2009 for an overview and discussion).

that present and future rights and obligations of the transaction partners are specified explicitly for all circumstances that might arise during and after the transaction. However, to design a contract covering all these contingencies is costly, thus reducing the gains from trade, or is even unfeasible. Thus, mutually beneficial economic exchange – cooperation – presupposes the solution of a social dilemma due to the problem of incomplete and implicit contracts (see Weber [1921] 1976: 409 for similar arguments in his sociology of law and Macaulay 1963 as a "modern classic").[2]

In this chapter, we discuss modeling approaches to social dilemmas and cooperation that use game-theoretic tools. Thus, we focus on approaches that take up Parsons' challenge and indeed try to specify conditions for cooperation of rational actors in social dilemmas. Such approaches avoid the "normative solution" proposed in various versions by Parsons and other classical sociologists, specifically in the functionalist tradition. According to the normative solution, cooperation is a result of internalized and shared norms and values as well as norm-conforming behavior. However, this approach only shifts the problem to the explanation of how such norms and values emerge and are maintained and is hardly compatible with the observation that the degree of norm-conformity varies not only between actors but varies also for the same actors over time (see also the chapter by Tutić et al. in this Handbook). Many game-theoretic approaches rather follow Coleman's (1964: 166-167) radical suggestion for taking up Parsons' challenge: "Hobbes took as problematic what most contemporary sociologists take as given: that a society can exist at all, despite the fact that individuals are born into it wholly self-concerned, and in fact remain largely self-concerned throughout their existence. Instead, sociologists have characteristically taken as their starting point a social system in which norms exist, and individuals are largely governed by those norms. Such a strategy views norms as the governors of social behavior, and thus neatly bypasses the difficult problem that Hobbes posed [...] I will proceed in precisely the opposite fashion [...] I will make an opposite error, but one which may prove more fruitful [...] I will start with an image of man as wholly free: unsocialized, entirely self-interested, not constrained by norms of a system, but only rationally calculating to further his own self interest."

We employ non-cooperative games (see also the chapter by Tutić in this Handbook) as models for social dilemmas. This is likewise a way of taking up Parsons' challenge. Binding agreements or binding unilateral commitments that are not explicitly modeled as results of actors' decisions are excluded in non-cooperative games. Technically speaking, such agreements and commitments have to be modeled as moves in the extensive form of the game. Thus, non-cooperative games model Hobbes state of nature. In this way, we avoid shifting the problem to the explanation of the emergence and maintenance of external enforcement of agreements and commitments. In fact, the availability of external enforcement presupposes that a social dilemma has been solved that is due to the fact that actors can benefit from such enforcement even

[2] One should not overlook that cooperation in a social dilemma, while beneficial for the actors directly involved, need not be beneficial for third parties. Members of the Mafia (e.g., Gambetta 1993) or cartel members (e.g., Stigler 1964) are involved in social dilemma-like interactions, with cooperation being beneficial for the members, while undermining cooperation in such contexts is desirable from a societal perspective.

if they have not contributed to the costs of providing enforcement. Assuming a non-cooperative game implies that external enforcement, if it enters the analysis, has itself to be explained in the first place.

Using game theory and thus the assumption of rational behavior as a tool for analyzing social dilemmas and cooperation has other advantages, too (see Diekmann & Lindenberg 2001: 2751-2752). First, it becomes clear why cooperation in social dilemmas is an interesting problem that requires careful analysis. Second, rigorous modeling with clearly specified assumptions and implications becomes feasible. Third, the implications include testable hypotheses and can thus be linked with systematic empirical research. Fourth, precisely because assumptions are clearly specified, it becomes feasible to modify and adapt them when empirical evidence reveals that assumptions are problematic (cf. Lindenberg's 1992 "method of decreasing abstraction").

The chapter proceeds with an overview of standard examples of "real life" social dilemmas and associated game-theoretic models. We then discuss mechanisms of cooperation for social dilemmas. The chapter concludes with a brief summary and some recommendations for further reading. Throughout, we aim at a sketch of key examples and ideas rather than comprehensiveness or technical details. We try to emphasize issues that are relatively neglected in other surveys of the field. Our focus, in line with the general topic of this Handbook, on formal tools such as game-theoretic models and on simulations is one but not the only respect in which our survey differs from others, since quite some surveys (e.g., Kollock 1998) do not use such tools explicitly. We also aim at highlighting how formal models are related to substantive social science theory. While a review of empirical research is not the focus of this chapter, we do exhibit empirical and testable implications of formal models.

2 Social dilemmas: examples

We use common notation for strategies and payoffs, with C_i denoting actor i's $(i = 1, ..., n)$ cooperative strategy and D_i his defective strategy. $D = D_1, ..., D_n$ is the strategy combination such that all actors defect, while $C = C_1, ..., C_n$ is the strategy combination such that all actors cooperate. We assume that D is a subgame perfect equilibrium with payoffs $U_i(D) = P_i$ for each actor, while cooperation is associated with payoffs $U_i(C) = R_i > P_i$. We assume that C is not only a Pareto-improvement compared to D but that C is also Pareto-optimal. Typically, C will not be an equilibrium. In general, we interpret payoffs as cardinal utilities.

While game theory does include the assumption that actors behave rationally, given their utilities and preferences, game theory *as such* does not include assumptions on whether an actor's utility depends exclusively on the actor's own material and possibly monetary outcomes. Neither does game theory *as such* involve assumptions on, for example, an actor's risk preferences. Thus, in principle, additional assumptions that an actor's utility (also) depends on the outcomes of an interaction for the other actor, on the fairness of outcomes, on "social orientations", etc. are consistent with using a game-theoretic framework. In fact, later on in this chapter, we will turn to models that do use such assumptions. For the time being, however, and in line with Coleman's approach, we do use the selfishness assumption and assume that an actor's utility depends exclusively on the actor's own material and possibly monetary outcomes ("utility = own money").

2.1 Social dilemmas with two actors

Trust Game, Investment Game, and Prisoner's Dilemma

For our examples, we distinguish between two- and n-actor social dilemmas. Social and economic exchange provides instructive examples of social dilemmas in interactions of two actors. Exchange may involve different kinds of social dilemmas, with different games lending themselves as formal models. First, consider exchange problems that result from one-sided incentives for opportunistic behavior that emerge when exchange is sequential. In social exchange (Blau 1964), Ego helps Alter today, trusting that Alter will help Ego tomorrow. If Alter indeed provides help tomorrow, both Ego and Alter are better off than without helping each other. However, Alter faces an incentive to benefit from Ego's help today without providing help himself tomorrow. Anticipating this, Ego might not provide help in the first place. In this example, cooperation means providing help while refusing to provide help is defection. In economic exchange between a buyer and a seller (e.g., Dasgupta 1988), the buyer may be insufficiently informed on the quality of a good and thus has to trust the seller that he[3] will sell a good product for a reasonable price. The seller can honor trust by indeed selling a good product for a reasonable price. Buyer and seller are then both better off compared to the situation without a transaction. Trust thus increases efficiency in economic exchange (Arrow 1974). However, the seller could also abuse trust by selling a bad product for the price of a good one, thus securing an extra profit. The buyer is then worse off than had he decided not to buy. In this example, cooperation by the buyer thus means that he enters into the exchange, while cooperation by the seller means that he sells a good product for a reasonable price. The buyer defects when he does not enter into the exchange, while the seller defects by selling a bad product for the price of a good one.[4] In both examples, only one actor has incentives for opportunistic behavior, i.e., behavior that impairs the partner by exploiting the partner's cooperation: Alter in the case of social exchange and the seller in the case of economic exchange. The other actor – Ego and, respectively, the buyer – can foresee this and may thus avoid entering into the exchange. Thus, defection by Alter or the seller is motivated by greed, while defection by Ego or the seller is defensive and motivated by fear.

[3] Throughout, we use male pronouns to facilitate readability and without intending any gender-bias.

[4] Of course, the seller might have other incentives for opportunistic behavior. For example, he might be tempted to secure an extra profit by serving another client first, thus reducing the buyer's gain from trade through a delivery delay. Also, the incentive problems for buyer and seller could be reversed with only the buyer having an incentive for opportunistic behavior such as delaying payment.

The *Trust Game* (Camerer & Weigelt 1988; Dasgupta 1988; Kreps 1990a; see also Coleman 1990: chap. 5; see also the chapters by Gautschi as well as by Abraham & Jungbauer-Gans in this Handbook), is a standard model for such exchange problems. The game (see Figure 1) involves a trustor (actor 1) and a trustee (actor 2). The game starts with a move of the trustor who can choose between placing trust (cooperation) or not placing trust (defection). If trust is not placed, the interaction ends and the trustor receives payoff P_1, while the trustee receives payoff P_2. If trust is placed, the trustee chooses between honoring (cooperation) and abusing trust (defection). If he honors trust, the payoffs for trustor and trustee are $R_i > P_i, i = 1, 2$. If trust is abused, the payoff for the trustor is $S_1 < P_1$, while the trustee receives $T_2 > R_2$.

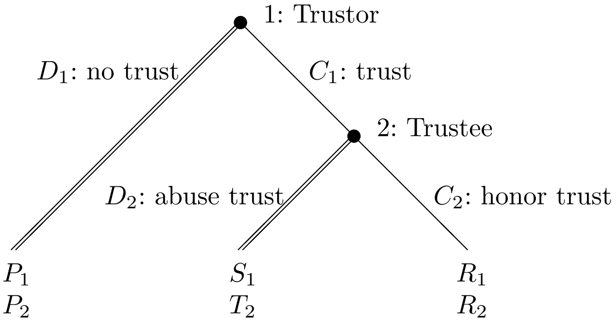

Fig. 1: The Trust Game ($S_1 < P_1 < R_1$, $P_2 < R_2 < T_2$); double lines indicate behavior in the unique subgame perfect equilibrium.

It is easily seen that the trustee's best-reply strategy against trust of the trustor is to abuse trust, while the trustor's best-reply strategy is then not to place trust. Not placing trust while placed trust would be abused is thus the unique subgame perfect equilibrium of the Trust Game and the solution of the game under the standard assumption that the solution has at least to be a subgame perfect equilibrium. Hence, the Trust Game is a social dilemma. Mutual cooperation (placing and honoring trust), while not an equilibrium, is Pareto-optimal and a Pareto-improvement compared to mutual defection (not placing trust, while trust would be abused) because $R_i > P_i$ for both actors. Note, too, that only actor 2 has an incentive for opportunistic behavior in the sense that he benefits from defection ($T_2 > R_2$). The trustor's best-reply strategy against cooperation of the trustee would be to cooperate herself by placing trust, since $R_1 > P_1$. The trustor's defection implies protection against the trustee's opportunism rather than an attempt to increase the trustor's own payoff by exploiting cooperation of the trustee.[5]

[5] It is useful to note that, allowing for mixed strategies, the Trust Game likewise has Nash equilibria in addition to the equilibrium in pure strategies of not placing trust, while placed trust would be abused. The set of Nash equilibria is the set of all strategy com-

Our next example of a social dilemma is a slightly more complex case of one-sided incentive problems in exchange that is closely related to Durkheim's limits of the contractual governance of economic transactions. We now assume that buyer and supplier do not make binary choices but have a larger set of feasible actions. For example, the buyer does not choose between "buying" and "not buying". Rather, he chooses how much time and effort the exchange partners allocate to writing an externally enforceable contract that reduces the seller's opportunities for exploiting the buyer but likewise reduces the gains from trade. Conversely, the seller chooses the degree to which he behaves opportunistically by not sharing these gains. If the buyer anticipates "much" opportunism of the seller, he may prefer an extensive but costly contract that reduces the sellers opportunities for exploiting the buyer. Both actors, however, would be better off without costly contracting and with larger and shared gains from trade.

The *Investment Game* (Berg et al. 1995; Ortmann et al. 2000) is a simple model for such situations. Again, actor 1 is the trustor and actor 2 is the trustee. The trustor can now choose the degree to which he trusts the trustee and the trustee can choose the degree to which he honors trust. More precisely, each actor has an endowment $E_i > 0$. The trustor chooses an amount e of her endowment to send to the trustee $(0 \leq e \leq E_1)$. Sending a larger e would mean that the buyer requires less extensive and thus less costly contractual safeguards. This "investment" e of the buyer is then multiplied by $m > 1$ and the trustee receives me. The parameter m represents how large the gains from trade are. Subsequently, the trustee chooses an amount g he returns to the trustor, with $0 \leq g \leq E_2 + me$. The seller thus decides on how to share the gains from trade. Afterwards, the game ends with the trustor receiving $U_1 = E_1 - e + g$ and the trustee receiving $U_2 = E_2 + me - g$. While e indicates how much the trustor trusts the trustee, g indicates how trustworthy the trustee is. It is easily seen that the Investment Game has a unique subgame perfect equilibrium: the trustee would never return anything, i.e., he would choose $g = 0$ for all e, while the trustor would send nothing so that $e = 0$. At the same time, this outcome of the game is Pareto-suboptimal: the game is a social dilemma since the actors forego all gains from trade. Thus, in the Investment Game, D_1 is the strategy to choose $e = 0$,

binations such that the trustor does not place trust, while the trustee would honor trust with probability $p \leq \frac{P_1 - S_1}{R_1 - S_1}$. Thus, in these equilibria, the probability of honoring trust is so small that the certain payoff P_1 associated with not placing trust is not smaller than the expected payoff $pR_1 + (1-p)S_1$ associated with placing trust if trust is honored with probability p. Therefore, the strategy of the trustor is a best reply against the mixed strategy of the trustee, while the mixed strategy of the trustee is likewise a best reply against the trustor's equilibrium strategy of not placing trust, since the trustee's payoff is then anyway P_2. However, all these equilibria are not subgame perfect, since abusing trust by the trustee is the unique equilibrium of the subgame that is reached after placement of trust by the trustor: a rational trustee would not be willing to make true his "promise" to honor trust with at least some small probability (namely, a probability not exceeding $\frac{P_1 - S_1}{R_1 - S_1}$), should the trustor place trust, since this would be inconsistent with payoff-maximization. Thus, the Trust Game is a nice example showing that subgame perfection can rule out certain "irrational" equilibria.

while D_2 is the strategy to choose $g = 0$ for all e. Hence, for the Investment Game, $U_i(D) = P_i = E_i$ for both actors.[6]

Cooperation can be conceptualized in different ways in the Investment Game, since the trustee could divide the sum of what he receives from the trustor (multiplied by m) plus his own endowment in many ways that imply Pareto-optimality as well as a Pareto-improvement compared to the outcome if the trustor sends nothing. For example, both actors are better off than in the equilibrium and the outcome is Pareto-optimal if the trustor sends everything and the trustee distributes the gains from trade equally, i.e., the trustor chooses $C_1 = E_1$ and C_2 implies that the trustee returns $g^* = E_1 + (m-1)\frac{E_1}{2}$ if the trustor sends his complete endowment. Then, $U_i(C) = R_i = E_i + (m-1)\frac{E_1}{2} > E_i$ for both actors. This conceptualization derives from using equality as a criterion for selecting a Pareto-optimal outcome. Another plausible conceptualization could use equity and distributive justice (e.g., Walster et al. 1978) as a criterion, i.e., to distribute the gains from trade so that the payoffs at the end of the game are proportional to the initial endowments. This would mean that the trustor chooses $C_1 = E_1$ and C_2 implies that the trustee chooses g^* so that $\frac{E_2+mE_1-g^*}{g^*} = \frac{E_2}{E_1}$ if the trustor sends his complete endowment. Then, $g^* = E_1\frac{E_2+mE_1}{E_1+E_2}$ and it is easily seen that $U_1(C) = R_1 = g^* > E_1$ and $U_2(C) = R_2 = E_2 + mE_1 - g^* = E_2\frac{E_2+mE_1}{E_1+E_2} > E_2$. Mutual cooperation is not an equilibrium in both cases but like in the Trust Game, only actor 2 has an incentive for defection in the Investment Game. The trustee's best-reply against C_1 is to return nothing so that $U_2(C_1, D_2) = E_2 + mE_1 > R_2$ for both conceptualizations of cooperation, while under both conceptualizations the trustor's best-reply strategy against C_2 is to cooperate himself by sending the complete endowment E_1. The trustor's equilibrium strategy D_1 to send nothing again implies exclusively protection against the trustee's opportunism.[7]

Incentive problems in exchange are often two-sided. For example, the seller has an incentive to sell a bad product for the price of a good one, while simultaneously the buyer has an incentive to delay payment. The *Prisoner's Dilemma*, presumably the most famous formal model for a social dilemma, can be used as a model for two-sided incentive problems in exchange with binary choices for buyer and seller (Hardin 1982). In the Prisoner's Dilemma, depicted in Table 1, actor 1 represents the row player, while

[6] One easily verifies that the Investment Game, similar to the Trust Game, has many Nash equilibria in pure strategies that are not subgame perfect, namely, all strategy combinations such that the trustor sends nothing, while the trustee's strategy is such that for all $e > 0$ he chooses some $g \leq e$. Nothing is sent in these equilibria but of course, a rational trustee, i.e., a trustee trying to maximize his payoff, would not return anything, would the trustor send something. There are also equilibria in mixed strategies that are not subgame perfect such that the trustee returns something with a sufficiently small probability. Again, subgame perfection rules out "irrational" equilibria.

[7] Note that the substantive interpretation of certain behaviors in the Investment Game in terms of "trust" and "abuse of trust" is somewhat problematic. If the amount e sent by the trustor is "small", returning a "small" amount g could also be interpreted as a punishment the trustee inflicts on the trustor for not trusting the trustee rather than as abusing trust.

actor 2 represents the column player. In this game, actors choose simultaneously[8] between cooperation and defection. In the Prisoner's Dilemma, mutual defection is not only the unique equilibrium but defection is also a dominant strategy for each actor: whatever the strategy of the other actor, own defection is always an actor's unique best reply. Other than in the Trust Game and in the Investment Game, *both* actors have an incentive to exploit cooperation of the other actor in the Prisoner's Dilemma, since $T_i > R_i$ for both actors. Thus, *both* actors are motivated by greed and fear. Note, too, that our terminology and notation for strategies and payoffs in social dilemma games is derived from meanwhile common terminology and notation for the Prisoner's Dilemma (e.g., Axelrod 1984).

Tab. 1: The Prisoner's Dilemma ($S_i < P_i < R_i < T_i$); the bold-faced payoffs indicate the unique equilibrium.

		Actor 2	
		Cooperation (C_2)	Defection(D_2)
Actor 1	Cooperation (C_1)	R_1, R_2	S_1, T_2
	Defection (D_1)	T_1, S_2	$\boldsymbol{P_1, P_2}$

There are other examples of two-actor games that model social dilemmas, with defection as the unique subgame perfect equilibrium. Such examples include the sequential Prisoner's Dilemma (actor j moves after actor i, knowing what i has chosen and thus being able to condition his own choice on the previous choice of i), the *Support Game* (Weesie 1988: 155–160; Hegselmann 1994; Vogt & Weesie 2004), which can be seen as a special case of the Prisoner's Dilemma, or Rosenthal's (1981) *Centipede Game*, of which the Trust Game is a special case.

The "theory dependence" of "social dilemma game"

Reviewing the literature on social dilemmas, it is striking that rather different typologies are available and that there seems to be no consensus for quite some games whether or not they should be considered as social dilemma games (see Van de Rijt & Macy 2009 for a detailed discussion). While it is straightforward to classify games with a unique subgame perfect equilibrium that is Pareto sub-optimal as social dilemma games, one should not overlook that it is far from obvious that *all* social dilemma games are of this type. This should not come as a surprise. After all, many games have more than one equilibrium, some of which may be Pareto-suboptimal, while others are not. It will thus depend on the "solution theory" (Harsanyi & Selten 1988) that

[8] "Simultaneous" in the sense that actor i, when making a choice, has no information on the choice of player j ($j \neq i$) and vice versa: no actor can condition his own behavior on the other actor's behavior. It is thus not necessary to assume that the actors decide *at the same time.*

is used for selecting one of the equilibria as the solution whether or not the game qualifies as a social dilemma game. In this sense, the definition of "social dilemma" becomes "theory dependent". Consequently, behavioral predictions for rational actors become theory dependent, too. This is an issue that arises already for well-known examples of 2 × 2-games such as those depicted in Table 2.

Tab. 2: A sample of 2 × 2 games; bold-faced payoffs indicate the equilibria in pure strategies.

		Actor 2	
		Left	Right
Actor 1	Top	**1, 1**	0, 0
	Down	0, 0	**1, 1**

a) Pure Coordination

		Actor 2	
		Left	Right
Actor 1	Top	**4, 4**	0, 2
	Down	2, 0	**3, 3**

b) Ranked Coordination

		Actor 2	
		Left	Right
Actor 1	Top	**4, 2**	0, 0
	Down	0, 0	**2, 4**

c) Battle of the Sexes

While actors have partly common and partly opposite interests in the Trust Game, the Investment Game, and the Prisoner's Dilemma, they have completely identical interests in *Pure Coordination* (Table 2a). A core feature of coordination problems is that actors, say, two co-authors, can choose between alternatives such as two different word processing programs or two different statistical software packages. The co-authors both benefit if they manage to coordinate so that they use the same software, while they both suffer if they do not succeed in coordinating their behavior. Pure Coordination 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. There is also a mixed equilibrium such that each actor chooses each of his two pure strategies with probability $\frac{1}{2}$. This equilibrium is Pareto-suboptimal and less attractive for each actor than the two equilibria in pure strategies. Payoff dominance is an often used criterion for selecting an equilibrium as the solution and according to this criterion, one of the two pure strategy equilibria would be selected and the game would not be a social dilemma game. However, this leaves the question open how actors – without communication and without repeated play – can coordinate their choices. A solution theory such as Harsanyi's (1977) thus selects the equilibrium in mixed strategies as the solution, since it satisfies – other than the two pure strategy equilibria – certain stability criteria. Under this theory, Pure Coordination *is* a social dilemma.

In *Ranked Coordination* (Table 2b), there are again two equilibria in pure strategies, (Top, Left) and (Down, Right). Again there is also an equilibrium in mixed strategies which is less interesting for our discussion. Both actors prefer (Top, Left) over (Down, Right). In terms of our example of software packages, both in Pure Coordination and Ranked Coordination, the co-authors can choose between two packages A and B. In Pure Coordination, it only matters whether they manage to use the same package and it does not matter whether that package happens to be A or B. In Ranked Coordination, however, both co-authors benefit even more if they manage to coordinate so that they choose one rather than the other package. Moreover, in Pure Coordination, actors are indifferent with respect to the two outcomes such that they fail to coordinate. In Ranked Coordination, no actor is indifferent between these outcomes and they have opposite interests with respect to these two outcomes. Thus, in the case of the two software packages, both co-authors favor the situation where they both use package A (say, Stata) over the situation where they both use B (say, SPSS). However, since other colleagues with whom they work on other joint papers, typically use package B, each of the two co-authors would prefer to use B in the situation such that coordination fails. In Ranked Coordination, (Top, Left) qualifies as the solution if payoff dominance is used as a criterion, while the solution would be (Down, Right) if risk dominance is used for equilibrium selection. Roughly, (Down, Right) fulfills the risk dominance criterion because Down is the more attractive strategy for actor 1 under the assumption that it is equally likely that actor 2 chooses Left or Right, while Right is the more attractive strategy for actor 2 under the assumption that it is equally likely that actor 1 chooses Top or Down. It thus again depends on the solution theory whether the game is or is not a social dilemma game. Note furthermore that both coordination games highlight that it need not be the case that cooperation is *not* an equilibrium: in both games, cooperation in the sense of successful tacit coordination on a Pareto-optimal strategy combination is also equilibrium behavior.[9]

[9] The literature offers quite some other games that model further variants of coordination games. Sometimes, authors use different labels for the same games. For example, a variant of Ranked Coordination is the Assurance Game (Sen 1974) which is also known as Stag Hunt (e.g., Van de Rijt & Macy 2009). See Camerer (2003: chap. 7) for an extended discussion of coordination games and a review of the literature.

Situations involving bargaining problems can have social dilemma characteristics, too. The *Battle of the Sexes*, depicted in Table 2c, is an example. The game models a bargaining problem such that (Top, Left) is attractive for actor 1, while (Down, Right) is attractive for actor 2. In this respect, the actors have conflicting interests. At the same time, the actors have a common interest in avoiding the conflict outcomes (Top, Right) and (Down, Left) that obtain when, for example, each actor insists on playing the strategy that is associated with his own most preferred outcome, given that the other actor concedes. Both (Top, Left) and (Down, Right) are equilibria and they are both Pareto-optimal. The game has a third equilibrium in mixed strategies, with each actor choosing each of his two pure strategies with probability $\frac{1}{2}$. This equilibrium is Pareto-suboptimal and it is indeed less attractive for both actors than the two equilibria in pure strategies. Thus, for actor 1 the equilibrium in mixed strategies is not only less attractive than his most preferred equilibrium (Top, Left) but it is also less attractive for actor 1 than the equilibrium (Down, Right), with actor 1 conceding to actor 2. Nevertheless, it is hard to see how to single out one of the two equilibria in pure strategies rather than the other as the solution of the game. Given the symmetry of the game ("it looks the same for each of the actors"), it seems plausible to require a symmetric solution ("each actor chooses the same strategy"). Indeed a solution theory such as Harsanyi's (1977) implies that the symmetric equilibrium in mixed strategies is the solution so that Battle of the Sexes would be a social dilemma.[10]

2.2 Social dilemmas with n actors

We now turn to social dilemmas with $n \geq 2$ actors. Cooperation in the sense of peacefully living together in Hobbes' state of nature is a social dilemma with more than two actors. Other examples are environmental public good problems (e.g., Hardin 1968) and Olson's (1965) case of the provision of public goods by organizations such as trade unions. We briefly discuss three important examples of game-theoretic models.

Taylor ([1976] 1987) has introduced an *n-actor Prisoner's Dilemma* that generalizes the two-actor version (see, e.g., Schelling 1978 for a more restrictive definition of the game). Each actor i can choose between cooperation (C_i) and defection (D_i). Hence, Taylor considers binary choice. Each actor's payoff U_i depends exclusively on his own choice and the number v ($0 \leq v \leq n-1$) of other actors j choosing C_j. Since the game is a Prisoner's Dilemma, it is reasonable to assume that D_i is a dominant strategy, i.e., $U_i(D_i, v) > U_i(C_i, v)$ for all v. Hence, each actor has an incentive to defect when some other actors cooperate. For a Prisoner's Dilemma, one likewise requires that each actor prefers cooperation of all actors to defection of all actors, i.e., $U_i(C_i, n-1) > U_i(D_i, 0)$. In addition, Taylor assumes that each actor's payoff when he himself defects and at least one other actor cooperates is larger than his payoff when everybody defects, i.e., $U_i(D_i, v) > U_i(D_i, 0)$ for all $v > 0$. Under these assumptions, defection by each actor is the unique equilibrium of the game. Moreover,

[10] Again, Battle of the Sexes is discussed under different labels in the literature. For example, Harsanyi (1977) uses "Bargaining Deadlock".

the equilibrium is Pareto-suboptimal. On the other hand, cooperation by each actor, while not an equilibrium, is a Pareto-improvement compared to universal defection and is also Pareto-optimal. Clearly, then, the n-actor Prisoner's Dilemma is a social dilemma game.

The *Public Goods Game* (e.g., Gächter & Thöni 2011) is a model for a social dilemma with n actors such that actors do not have to make a binary choice between cooperation and defection. Rather, they can choose between more or less cooperation. Each actor has an endowment E. Actors simultaneously choose a contribution g_i to the public good with $0 \leq g_i \leq E$. The total amount contributed, i.e., $g = \sum g_i$, is multiplied by m, with $1 < m < n$, and subsequently mg is divided equally among the actors. Hence, each actor's payoff U_i depends on his own contribution and the contribution of all other actors, with $U_i = E - g_i + \frac{m}{n}g$. Since $m < n$, the individual return from i's own contribution is smaller than the individual contribution itself ($\frac{m}{n}g_i < g_i$). Thus, contributing nothing is a dominant strategy for each actor and the game has a unique equilibrium with $U_i(D) = E$ for each actor and D denoting defection ($g_i = 0$) by each actor. Obviously, this outcome is Pareto-suboptimal: since $m > 1$, each actor is better off and the outcome is Pareto-optimal when each actor contributes his whole endowment. Then, each actor's payoff is $U_i(C) = mE > E$, with C denoting cooperation ($g_i = E$) by each actor. This shows that the Public Goods Game is a social dilemma game with each actor having an incentive to defect.[11]

Our final and meanwhile well-known example of a social dilemma with n actors is Diekmann's (1985) *Volunteer's Dilemma*. Actors have binary choices. They decide simultaneously whether or not to provide a collective good. The good is costly and will be provided if at least one actor – the "volunteer" – decides to provide. Contributions by more than one actor are feasible and then each actor pays the full costs of providing the good but contributions of more than one actor do not affect the utility level of any actor. The new feature compared to the n-actor Prisoner's Dilemma and the Public Goods Game is that the costs of providing the collective good are smaller than the gains from the good. Thus, the matrix in Table 3 summarizes the normal form of the game, with the rows representing an actor's strategies, namely, to provide the good (PROV) or not to provide (DON'T), with columns indicating the number of other actors who choose PROV, and cells representing the actor's payoffs as a function of his own strategy and the number of other actors who choose PROV.

In the Volunteer's Dilemma, thus, each actor has an incentive to provide the public good when nobody else is providing, while all other actors have an incentive not to provide if there is at least one volunteer. Diekmann (1985) discusses the bystander intervention and diffusion of responsibility problem (Darley & Latané 1968) as an example of a social situation for which the Volunteer's Dilemma is a reasonable model. This is a situation with n actors witnessing an accident or a crime. Everybody would feel relieved if at least one actor would help the victim by, for example, calling the

[11] Note that the two-actor Public Goods Game relates to the standard Prisoner's Dilemma with two actors much in the same way as the Investment Game relates to the Trust Game. The Investment Game is a variant of the Trust Game and the two-actor Public Goods Game is a variant of the Prisoner's Dilemma with continuous rather than binary choices for the actors.

Tab. 3: The Volunteer's Dilemma ($R > K > 0$; $n \geq 2$).

	Number of other actors choosing PROV				
	0	1	2	...	$n-1$
PROV	$R-K$	$R-K$	$R-K$	...	$R-K$
DON'T	0	R	R	...	R

police. However, providing help is costly and each actor might be inclined to abstain from helping, hoping that someone else will help.

The Volunteer's Dilemma has n equilibria in pure strategies. These are the strategy combinations with exactly one volunteer choosing PROV while all other actors choose DON'T. Each of these equilibria is Pareto-optimal. However, the equilibria involve a bargaining problem, since each actor prefers the equilibria with another actor as the volunteer to the equilibrium where he himself is the volunteer. Moreover, while the game is symmetric, the n equilibria in pure strategies require that actors do not behave the same. It can be shown that the Volunteer's Dilemma has a unique symmetric equilibrium in mixed strategies such that each actor chooses DON'T with probability $p^* = (\frac{K}{R})^{\frac{1}{n-1}}$. It follows from elementary properties of equilibria in weak strategies that each actor's expected payoff associated with the symmetric equilibrium in mixed strategies is $R - K$. Hence, each of the equilibria in pure strategies is a weak Pareto-improvement since in each of those equilibria the volunteer is not worse off, while all other actors are better off.

From a game-theoretic perspective, the symmetric equilibrium in mixed strategies is a plausible candidate for the solution of the Volunteer's Dilemma. An important feature of the Volunteer's Dilemma is that indeed another combination of mixed strategies exists that is symmetric so that each actor chooses the same strategy, that is Pareto-optimal, and is a (strong) Pareto-improvement because it is associated with expected payoffs that are higher for each actor than his expected payoff in the symmetric mixed strategy equilibrium. Namely, if each actor chooses DON'T with probability $p^{**} = (\frac{K}{nR})^{\frac{1}{n-1}}$, each actor's expected payoff is $R - qK$ with $1 > q = 1 - p^*(1 - \frac{1}{n}) > \frac{1}{n}$. Since the strategy combination such that each actor chooses DON'T with probability p^{**} is not an equilibrium, it follows that the Volunteer's Dilemma is indeed a social dilemma game. The mixed strategy of choosing DON'T with probability p^* represents defection, while cooperation would imply that each actor chooses DON'T with probability p^{**}. Note that $p^* > p^{**}$, i.e., cooperation in the Volunteer's Dilemma implies a larger individual probability of providing the public good than defection.

Olson's (1965) classic contribution focused on the problem of group size effects on the production of collective goods. Group size effects on cooperation in social dilemmas are a very similar problem (see Raub 1988). Also, the problem of how group size affects public good provision and cooperation in social dilemmas is a paradigmatic example of a micro-macro problem in social science. After all, group size is an example of a macro-condition that might affect individual behavior on the micro level, while "provision of a public good" or a "Pareto-optimal outcome in a social dilemma problem" are examples of macro-effects of individual behavior. For the Volunteer's Dilemma, a natural way to analyze this issue is to consider how an increase in the number of actors affects the probability of the provision of the public good. Under our assumptions on individual behavior, group size affects public good provision in the Volunteer's Dilemma through two different mechanisms. First, since it is sufficient for the good to be provided that one single actor is willing to bear the costs and since all actors cooperate with positive probability in the symmetric mixed equilibrium, there is a positive effect of increasing group size since the number of actors increases who may decide to provide the good. Second, there is a negative effect of increasing group size, since each actor's individual probability to provide the good decreases with increasing n. What is the total effect? For the symmetric mixed equilibrium, the second effect outweighs the first, since the probability that the collective good will be provided, i.e., that there is at least one volunteer, is $1-(p^*)^n = 1-(\frac{K}{R})^{\frac{n}{n-1}}$. Hence, the probability that the public good is provided decreases with increasing n. It is interesting to note that the opposite result is obtained for the group size effect on public good production when actors cooperate by choosing the mixed strategy not to provide the public good with probability p^{**}. In that case, $1-(p^{**})^n = 1-(\frac{K}{nR})^{\frac{n}{n-1}}$ is the probability that the good will be provided and this probability increases with increasing n. We will briefly return to the issue of group size effects on cooperation in a social dilemma in the next section.

3 Mechanisms of cooperation

The literature meanwhile provides a sizeable number of mechanisms that can induce cooperation in social dilemmas (see Kollock 1998 for an overview). Quite some game-theoretic modeling focuses on how the "embeddedness" (Granovetter 1985) of a social dilemma affects cooperation. Weesie & Raub (1996) distinguish between embeddedness in ongoing relations of the actors, in networks of relations, and in institutions. This distinction has meanwhile become rather common (see Diekmann & Lindenberg 2001). Such an approach is interesting from a substantive social science perspective since it is in line with Coleman's (1987) heuristic advice to combine robust assumptions on rational behavior with more complex assumptions on social structure as well as with Granovetter's programmatic sketch (1985) that likewise advocated precisely such a combination of assumptions. We sketch how ongoing relations and institutional embeddedness can induce cooperation (see our chapter on social networks in this Handbook for effects of network embeddedness). For our sketch, we use the Trust Game as a convenient example, while much of the analysis is applicable to a broad class of other social dilemma games as well.[12]

[12] Since we focus on the Trust Game as a paradigmatic example of social dilemma games, we do not cover the literature that focuses specifically on mechanisms of cooperation in social dilemmas with n actors. Ostrom (1990) is an example of an influential study, while Ledyard (1995) provides an overview. See also Balliet et al. (2011).

3.1 Conditional cooperation in the repeated Trust Game

Assume that a buyer purchases repeatedly from the same seller and that there are one-sided incentives for opportunistic behavior of the seller in each transaction. A reasonable formal model for such a situation is a repeated Trust Game (e.g., Kreps 1990a). More precisely, the Trust Game is played indefinitely often in rounds $1, 2, \ldots, t, \ldots$ so that after each round t, another round $t+1$ is played with probability w $(0 < w < 1)$, while the repeated game ends after each round with probability $1 - w$. The focal Trust Game is thus embedded in a more complex game. In each round, trustor and trustee observe each other's behavior. In the repeated game, a strategy is a rule that prescribes an actor's behavior in each round t as a function of the behavior of both actors in the previous rounds. An actor's expected payoff for the indefinitely repeated Trust Game is the discounted sum of the actor's payoffs in each round, with the continuation probability w as discount parameter. For example, a trustor who places trust throughout the repeated game, with trust being honored throughout, receives payoff $R_1 + wR_1 + \ldots + w^{t-1}R_1 + \ldots = \frac{R_1}{1-w}$. Thus, using Axelrod's (1984) well-known label, the continuation probability w represents the "shadow of the future": the larger w, the more an actor's payoff from the repeated game depends on what the actor receives in future rounds.

In the indefinitely repeated Trust Game, the trustor can use a conditional strategy that rewards a trustee who honors trust in a focal Trust Game by placing trust again in future games. Conversely, a conditional strategy of the trustor can punish abuse of trust by the trustee in the focal Trust Game through not placing trust in at least some future games. Other forms of rewards and punishment are excluded in this simple scenario.

If the trustor uses reciprocity (Gouldner 1960; Blau 1964; Diekmann 2004; see also the chapter by Berger & Rauhut in this Handbook) in the sense of implementing a conditional strategy, the trustee can gain T_2 rather than R_2 in the current Trust Game by abusing trust. However, abusing trust will then be associated with obtaining only P_2 in (some) future encounters with no trust placed by the trustor, while honoring trust will result in larger payoffs than P_2 in those future encounters if the trustor goes on placing trust. Moreover, the larger the shadow of the future is, the more important are the long-term effects of present behavior. Thus, anticipating that the trustor may use a conditional strategy, the trustee has to balance short-term $(T_2 - R_2)$ and long-term $(R_2 - P_2)$ incentives. It can be shown that conditional cooperation (Taylor [1976] 1987) can be a basis for rational trust in the sense that the indefinitely repeated Trust Game has an equilibrium such that trust is placed and honored in each round. Cooperation in the Trust Game is then driven exclusively by long-term, "enlightened" self-interest of the actors ("cooperation of rational egoists").[13]

[13] For simplicity, we only consider the feasibility of conditional cooperation as an effect of ongoing interactions. Another effect of ongoing interactions is that actors can learn about unobservable characteristics of their partners through such interactions. For example, a buyer may be able to infer from the outcomes of previous transactions whether the seller is not only trustworthy in the sense of abstaining from opportunistic behavior but is also competent enough to deliver reasonable quality. See Buskens & Raub (2002) for

Consider the strategy of the trustor that is associated with the largest rewards for trustworthy behavior of the trustee and with the most severe sanctions for untrustworthy behavior. This is the strategy that prescribes to place trust in the first round and also in future rounds, as long as trust has been placed and honored in all previous rounds. However, as soon as trust is not placed or abused in some round, the trustor refuses to place trust in any future round. Such a strategy is often labeled a "*trigger strategy*" because deviation of the trustee from the "prescribed" pattern of behavior triggers a change in the trustor's behavior. Straightforward analysis shows (see, e.g., Friedman 1990) that always honoring trust (and always abusing trust as soon as there has been any deviation from the pattern "place and honor trust") is a best reply of the trustee against a trigger strategy of the trustor if and only if

$$w \geq \frac{T_2 - R_2}{T_2 - P_2}. \tag{1}$$

This condition requires that the shadow of the future is large enough compared to $\frac{T_2 - R_2}{T_2 - P_2}$, a convenient measure for the trustee's temptation to abuse trust in a repeated Trust Game.

If condition 1 applies, the indefinitely repeated Trust Game has an equilibrium such that trust is always placed and honored. This equilibrium is likewise subgame perfect.[14] This implies that the trustor's (implicit) promise to reward trustworthy behavior of the trustee by placing trust again in the future and her (implicit) threat to punish abuse of trust by not placing trust again are credible. Enlightened self-interest can thus be a basis for trust among rational and selfish actors in the sense of placing and honoring trust being equilibrium behavior.[15] The equilibrium, however, is not unique. For example, never placing trust, while placed trust would always be abused is always an equilibrium of the indefinitely repeated game. The "folk theorem" (Rasmusen 2007: chap. 5.2) for repeated games implies that the indefinitely repeated Trust Game has many other equilibria, too, for large enough w. Thus, an equilibrium selection problem emerges. A typical, though sometimes implicit, argument in the literature on equilibrium selection in this context is payoff dominance. In the indefinitely repeated Trust Game, an equilibrium that implies placed and honored trust throughout the game is evidently not payoff dominated by other equilibria, while such an equilibrium payoff-dominates the no-trust-throughout equilibrium.

a systematic discussion of embeddedness effects that distinguishes between conditional cooperation ("control") and learning as mechanisms that can support cooperation.

[14] This shows that rational actors may cooperate in a non-cooperative game. The trigger strategy is not the only conditional strategy that can be used to stabilize trust and trustworthiness as a result of equilibrium behavior of rational actors. Other conditional strategies that use less severe punishments than the trigger strategy can do so, too. However, one then needs further equilibrium conditions rather than exclusively condition 1.

[15] Coleman clearly intuited this result when he argued that an important feature of socialization is "coming to see the long-term consequences to oneself of particular strategies of action" rather than the internalization of norms (1964: 180). Voss (1982) seems to be the first sociologist who realized explicitly that the theory of repeated games has important implications for the problem of order and cooperation in social dilemmas.

It can be argued that the trigger strategy equilibrium for the indefinitely repeated Trust Game is implausible. For example, the equilibrium implies that trust is *always* placed and honored, while one might rather expect less than "perfect" trust, even under favorable conditions for trust. One can show (see, e.g., Taylor [1976] 1987) that there are also other equilibria that induce placement of trust and honoring trust only in some, rather than in all rounds of the game, while this pattern of behavior is again backed by a variant of the trigger strategy: as soon as there is a deviation from the "prescribed" pattern, the trustor never places trust again. Then, however, the problem becomes even more severe to select one out of a wealth of such equilibria as a "solution candidate". Moreover, although no deviations occur in equilibrium, it may seem implausible that trust *would* break down *completely* after the first deviation. A complete breakdown of all future trust may seem implausible for a single deviation from the trigger strategy equilibrium in the sense of honored trust in each round of the game as well as for a single deviation from other patterns of behavior that are backed by a variant of a trigger strategy. This counter-intuitive feature can be circumvented, for example, by considering a game with imperfect monitoring (e.g., Green & Porter 1984). Assume that the trustor, after placing trust, cannot observe the trustee's behavior, but can only observe the outcome of that behavior. This outcome, in turns, depends on the trustee's behavior but also on chance: a low payoff for the trustor after placement of trust can be due to abuse of trust by the trustee but can also be due to "bad luck". Such a scenario is much more difficult to analyze. The trustor now has to solve an "optimal punishment" problem. If the trustor never punishes, or applies too lenient punishments, a rational trustee would always abuse trust. But too severe punishments imply more than necessary (in terms of deterring the trustee from abusing trust) losses for trustor and trustee. Equilibrium behavior that generates some honored trust throughout the game now requires that the trustor punishes the trustee by placing no trust occasionally rather than eternally.

Even though the trigger strategy equilibrium for the indefinitely repeated Trust Game may have implausible features, consider an interpretation of the equilibrium condition that follows a widely used – while often left implicit – logic for deriving testable hypotheses from game-theoretic models (see Buskens & Raub 2013 for discussion). Rather than claiming that actors indeed use trigger strategies, one proceeds from the observation that condition 1 is a necessary and sufficient condition for equilibria in the indefinitely repeated Trust Game such that trust is placed and honored throughout the game. One then assumes that placing and honoring trust becomes more likely when the condition becomes less restrictive. This leads directly to testable hypotheses on effects of embeddedness. Specifically, one would expect that the likelihood of placing and honoring trust increases in the shadow of the future w and decreases in the temptation $\frac{T_2-R_2}{T_2-P_2}$ for the trustee.

The results for the indefinitely repeated Trust Game can be generalized. For example, analogous results hold for an indefinitely repeated Investment Game. Friedman (1971, 1990) shows that analogous results apply to a broad class of indefinitely repeated two- and n-actor games. Roughly speaking, if a social dilemma game is repeated indefinitely often and the shadow of the future is large enough relative to the short-term incentives of the actors, there exists an equilibrium of the indefinitely repeated game such that the actors cooperate: the equilibrium of the repeated game induces a Pareto-optimal outcome and a Pareto-improvement compared to the Pareto-suboptimal solution of the original dilemma. Also, for repeated social dilemma games with n actors, one can study the group size effect on cooperation by a comparative statics analysis of whether an appropriate generalization of the equilibrium condition 1 becomes more or less severe when n increases (see Raub 1988). Of course, these generalizations should be interpreted with care. For example, trigger strategies require the observability of the behavior of other actors. Hence, the underlying assumption that each actor receives reliable information on each other actor's behavior in each round of the game is crucial, while such an assumption will often be rather problematic from an empirical perspective in games with many actors (see, e.g., Bendor & Mookherjee 1987).

A Trust Game is embedded in a network when, for example, the trustee is involved in interactions with other actors with ties among each other as well as with the trustor so that information can be exchanged about the trustee and his behavior. Next to direct reciprocity exercised by the trustor who interacts with the trustee in the focal Trust Game, network embeddedness allows for indirect reciprocity exercised by other partners of the trustee. A trustee contemplating to honor or abuse trust in a focal Trust Game now has to consider not only future sanctions by the trustor with whom he interacts in the focal Trust Game but also sanctions that can be applied by other future interaction partners who receive information on the trustee's behavior in the focal Trust Game and who may condition their future behavior on that information. Network embeddedness and cooperation in social dilemmas will be discussed in more detail in the chapter on social networks in this Handbook.

3.2 Cooperation through institutional embeddedness

Institutions often enhance embeddedness by allowing actors to inform others, thus enhancing opportunities to cooperate conditionally. Modern examples are eBay's feedback forum and similar reputation systems used in the Internet economy. An institution such as the eBay feedback forum allows buyers to evaluate sellers and to collect information on sellers from other buyers. Similarly, sellers can provide and receive feedback on buyers. Fascinating cases of similar institutions in medieval trade are the Maghribi traders' coalition (see Greif 2006 for a comprehensive treatment) and the law merchants (Milgrom et al. 1990; see also Klein 1997 for more examples and Schramm & Taube 2003 for the more recent example of the Islamic *hawala* financial system). While they help actors to overcome social dilemmas in economic exchange, such institutions cannot be taken for granted, for example, due to incentive problems associated with the provision of (correct) information and feedback. Hence, a strong feature of the models provided by Greif as well as Milgrom et al. is that the institutions are "endogenized" in the sense that it is shown that they are themselves result of equilibrium behavior in repeated games.

Institutions can help actors also in other ways in overcoming social dilemmas. Contract law and other institutions often provide opportunities for actors to modify themselves their own (future) incentives or, as Coleman (1990) put it, to construct their social environment. Actors do so by incurring *commitments* (Schelling 1960; Williamson 1985). For example, a seller in the role of the trustee voluntarily provides a guarantee before the Trust Game itself is played. The guarantee modifies the subsequent incentives for trustor and trustee in the Trust Game. Commitments such as guarantees can promote trust by reducing the trustee's incentive for abusing trust, by providing compensation for the trustor in case trust is abused, or by signaling that the trustee will not (or cannot) abuse trust. Game-theoretic models can be used to specify conditions such that commitments are incurred and induce placing and honoring trust (e.g., Weesie & Raub 1996; Raub 2004). These models allow for deriving hypotheses on how characteristics of the commitment such as the costs associated with incurring a commitment, the size of the reduction of the trustee's incentive to abuse trust, or the size of the compensation for the trustor in case of abused trust affect the likelihood of incurring a commitment as well as the likelihood of placing and honoring trust. In these models, a context that provides opportunities for incurring a commitment is assumed as exogenous. The commitment itself can then be conceived as a "private institution", voluntarily created by the actors involved in a social dilemma for overcoming the dilemma. A strength of the models is then again that the private institution is not taken for granted but is itself an outcome of equilibrium behavior.

Institutional embeddedness can be a substitute as well as a complement for embeddedness in the sense of ongoing interactions or network embeddedness. Given institutional embeddedness, actors can overcome trust problems and other social dilemmas even if ongoing interactions or networks are absent or are insufficient to promote trust, for example, due to large incentives for abusing trust ("golden opportunities"). Also, some models are meanwhile available that study effects of embeddednes in ongoing interactions, network embeddedness, and institutional embeddedness simultaneously (e.g., Weesie et al. 1998).

4 Social preferences: cooperation in one-shot social dilemmas

Until now, we have studied how embeddedness of a social dilemma can induce cooperation of rational and selfish actors. The rationality assumption is captured in the notion of game-theoretic equilibrium behavior. The selfishness assumption is an assumption on properties of actors' utility and on how physical and psychic outcomes of interactions are "converted" into utility. More specifically, the selfishness assumption entails that an actor's utility depends exclusively on his own material and possibly monetary outcomes ("utility = own money").

A one-shot social dilemma is an "isolated encounter" for the actors in the sense that they cannot condition behavior in future interactions on what happens in the dilemma and that they cannot modify the incentive structure of the dilemma through, for example, incurring commitments ex ante. Then, by definition, rational and selfish actors will defect rather than cooperate. Isolated encounters are hardly a standard feature of interactions in social and economic life. Rather, one-shot social dilemmas are typically studied in the laboratory, using one-shot social dilemma games that are well-suited to explore alternatives for the selfishness assumption, because other factors such as embeddedness characteristics can be controlled.

Once again, we use the Trust Game from Figure 1 as an example, now assuming that the payoffs represent monetary incentives or points converted into money at the end of the experiment. Assuming rational and selfish behavior, the prediction for the one-shot Trust Game is no trust and, if trust would be placed anyway, it would be abused. This prediction is clearly rejected (see Snijders 1996; Snijders & Keren 1999, 2001). Similar results are found for other social dilemmas such as the Investment Game (Berg et al. 1995; see Camerer 2003: chap. 2.7 for an extensive review), the standard Prisoner's Dilemmas (e.g., Sally 1995) and n-person dilemmas (e.g., Ledyard 1995). Thus, the experimental evidence indicates that opportunism is not ubiquitous in one-shot social dilemmas (see the chapter by Pointner & Franzen in this Handbook for further review and references).

Different game-theoretic approaches have been suggested that account for the empirical regularities in experiments with one-shot social dilemma games (see Fehr & Schmidt 2006 for an instructive overview). First, one could relax the rationality assumption and employ a *bounded rationality* perspective. For example, one could assume that subjects are used to repeated interactions in life outside the laboratory. As we have seen, cooperative behavior can be a result of equilibrium behavior in repeated social dilemmas. The assumption then is that subjects erroneously apply rules in one-shot interactions that are appropriate when interactions are repeated (see, e.g., Binmore 1998). More generally, Binmore (1998: chap. 0.4.2) argues that behavior in experimental games can be expected to be consistent with the assumption of selfish game-theoretic rationality only if the game is easy to understand, adequate incentives are provided, and sufficient time is available for trial-and-error learning (see Kreps 1990b for similar arguments).

Second, there are approaches that maintain the rationality assumption but modify the selfishness assumption (prominent examples include Fehr & Schmidt 1999 and Bolton & Ockenfels 2000; see Fehr & Schmidt 2006 and Fehr & Gintis 2007 for overviews). These approaches thus abandon the assumption that subjects care exclusively about their own material resources. Rather, it is assumed that some subjects, have *social preferences*. It is quite often argued (e.g., Fehr & Gintis 2007) that such preferences are due to socialization processes and internalized social norms and values. Also, subjects differ with respect to their social preferences: there are selfish subjects as well as subjects with social preferences. Finally, subjects are incompletely informed on the preferences of other subjects.[16]

To get a flavor of models employing assumptions on social preferences, consider Snijders' (1996; see also Snijders & Keren 1999, 2001) *guilt model*, a simplified version

[16] See the chapters by Pointner & Franzen on fairness and by Gautschi on information and signals in this Handbook for further information on models using fairness assumptions and on how behavior in one-shot and repeated interactions is affected by incomplete information and signaling. See Lindenberg's (e.g., 2001) theory of "social rationality" for an interesting alternative approach to explaining behavior in social dilemmas and other interaction situations that is hard to reconcile with the assumptions of rational and selfish behavior.

of the Fehr-Schmidt (1999) model of inequity aversion. Assume that actor i's utility is given by $U_i(x_i, x_j) = x_i - \beta_i \max(x_i - x_j, 0)$ with monetary payoffs x_i and x_j for the actors i and j and $\beta_i \geq 0$ a parameter representing i's guilt due to an inequitable allocation of monetary payoffs. Hence, in a Trust Game with payoffs in terms of money and $P_1 = P_2$ and $R_1 = R_2$, the trustee's utility from abused trust would be $T_2 - \beta_2(T_2 - S_1)$, while utilities correspond to own monetary payoffs in all other cases.[17] Furthermore, assume actor heterogeneity with respect to the guilt parameter β_i in the sense that there are actors with a large guilt parameter, while β_i is small or even equals zero for other actors, namely, those with selfish preferences. Finally, assume incomplete information of the trustor on the trustee's guilt parameter, with π being the probability that β_2 is "large enough" so that the trustee's utility from abusing trust is smaller than his utility from honoring trust, i.e., $T_2 - \beta_2(T_2 - S_1) < R_2$. "Large enough" thus means $\beta_2 > \frac{T_2 - R_2}{T_2 - S_1}$, with $\frac{T_2 - R_2}{T_2 - S_1}$ as a convenient measure of the temptation of an inequity averse trustee to abuse trust. Equilibrium behavior now requires that a trustee with $\beta_2 > \frac{T_2 - R_2}{T_2 - S_1}$ honors trust, while a trustor places trust if $\pi > \frac{P_1 - S_1}{R_1 - S_1}$. Again assuming that a certain equilibrium behavior becomes more likely when the condition for the existence of the equilibrium is less restrictive, one predicts that placing trust becomes more likely when the condition $\pi > \frac{P_1 - S_1}{R_1 - S_1}$ becomes less restrictive. Similarly, honoring trust should become more likely when the condition $\beta_2 > \frac{T_2 - R_2}{T_2 - S_1}$ becomes less restrictive. Furthermore, one could assume that π depends on the trustee's incentives and hence decreases in $\frac{T_2 - R_2}{T_2 - S_1}$. It follows from this model that the likelihood of placing trust decreases in the trustor's risk $\frac{P_1 - S_1}{R_1 - S_1}$ as well as in the trustee's temptation $\frac{T_2 - R_2}{T_2 - S_1}$ and that the likelihood of honoring trust decreases in the trustee's temptation $\frac{T_2 - R_2}{T_2 - S_1}$. These implications nicely correspond with experimental evidence (see Snijders 1996; Snijders & Keren 1999, 2001).

Obviously, assumptions on social preferences should be used with care (see, e.g., Camerer 2003: 101; Fehr & Schmidt 2006: 618): (almost) all behavior can be "explained" by assuming the "right" preferences and adjusting the utility function. Thus, one would prefer first of all parsimonious assumptions on social preferences, adding as few new parameters as possible to the model. Second, when assumptions on social preferences are employed, one should aim at using the same set of assumptions for explaining behavior in a broad range of different experimental games. Third, one should account not only for well-known empirical regularities but also aim at deriving and testing new predictions. It is therefore important from a methodological perspective that the same set of assumptions on social preferences is consistent not only with empirical regularities of behavior in Trust Games but also in other social dilemmas, in games involving distribution problems such as the Ultimatum Game (Güth et al. 1982; see Camerer 2003 for a survey; see also the chapter by Rieck in this Handbook) or the Dictator Game (Kahneman et al. 1986; see Camerer 2003 for a survey), and in market games. Fehr & Schmidt (1999; see Bolton & Ockenfels 2000 for similar

[17] Snijders thus neglects that the trustor may derive an additional disutility from abused trust because he envies the inequitable distribution. The Fehr-Schmidt model takes such a disutility into account.

arguments) argue that their model of inequity aversion succeeds in accounting not only for empirical regularities in a broad class of experimental games but is also consistent with selfish behavior in some settings and non-selfish behavior in others. This is due to heterogeneity between the actors with respect to their inequity aversion. The interaction between actors who are selfish and actors with (stronger) inequity aversion in a setting with incomplete information on other actors' preferences can be a driving force in inducing selfish behavior in settings such as experimental markets and quite some non-selfish behavior in, for example, social dilemma games. Finally, it becomes important to empirically discriminate between different assumptions on social preferences, using careful experimental designs that allow disentangling the different mechanisms assumed in different models of social preferences (see again Fehr & Schmidt 2006 for a survey).

Assumptions on social preferences cannot only account for cooperation in one-shot social dilemma games but also for related findings. For example, a meanwhile extensive literature shows experimental evidence that cooperation in social dilemma games can be enhanced when subjects can punish defection of other subjects, even if punishment is costly (see, e.g., Fehr & Gächter 2000 for important and early experimental work and Fehr & Gintis 2007 for an overview). Assuming selfish rationality, one would predict that actors would not use a costly punishment option and that such an option would not affect behavior in the dilemma game. Assumptions on social preferences can help accounting for the fact that punishment is used and that cooperation is affected by punishment options.

5 Simulation studies

This chapter focuses on game-theoretic models of social dilemmas and cooperation. Studies such as Ullmann-Margalit (1977) and specifically Taylor ([1976] 1987) pioneered the systematic application of game-theoretic tools in this field. Simulation studies have emerged as an important complementary tool. When models are no longer analytically tractable, simulation studies can be used as an alternative. Problems with analytical tractability can come from various sources. These typically originate from attempts to make the assumptions of highly simplified game-theoretic models more complex and presumably more realistic, in the spirit of Lindenberg's (1992) "method of decreasing abstraction". Related to the extensions of standard assumptions that are also discussed above, assumptions that increase the complexity of the game-theoretic models include assumptions on actor's rationality or "bounded" rationality, on the embeddedness of the actors and their interaction, on the complexity of games actors are involved in, on heterogeneity of actors, and on the evolutionary dynamics of the behavior. It is beyond the scope of this chapter to discuss simulation models on these issues in detail. Also, such a discussion would create considerable overlap with other chapters in this Handbook (see the chapter by Opp as well as the chapter by Flache & Mäs). Therefore, we restrict ourselves to some general remarks and links to further literature.

In the field of social dilemma research, simulation methods have been pioneered by Axelrod (1984). His seminal study involved a “tournament” in which a large variety of strategies for the repeated Prisoner’s Dilemma were pitched against each other, famously showing that Tit-for-Tat was the most successful strategy. In fact, Axelrod relaxed the assumption that actors are rational and implement equilibrium behavior. Rather, he assumed that actors have prescribed rules of play, including conditionally cooperative rules. What really highlights the added value of Axelrod’s simulation approach, is that he was also able to demonstrate that Tit-for-Tat was also the most successful strategy in an evolutionary context in which more successful strategies reproduce more. Later studies have qualified and extended these results (e.g., Nowak 2006). By now, evolutionary game theory (Weibull 1995) has become a major subfield of game theory that applies analytic modeling in combination with simulation techniques (see also the chapter by Amann in this Handbook).

An alternative approach to modeling bounded rationality is that actors are not forward looking in the sense of anticipating possible consequences of their behavior in repeated games, but are instead *backward* looking and adapt their behavior based on previous experiences. Among others, Macy (1991) and Flache & Macy (2002) provide examples of how hypotheses can be derived for the likelihood of the emergence of cooperative behavior in social dilemmas from such backward looking models. These models can be considered as special cases of a much more general class of agent-based models in which agents in interdependent situations can be modeled using relatively simple decision rules for the agents (see, e.g., Epstein 2006; Gilbert 2010; Squazzoni 2012; and the chapter by Flache & Mäs in this Handbook) without assuming that the actors are rational and thus employing equilibrium strategies.

An example in which the interaction structure becomes too complex to derive many interesting hypotheses is the model by Buskens (2002: chap. 3) in which trustors embedded in a network that allows for information transmission play Trust Games with one trustee. Buskens uses simulation to enumerate the equilibrium behavior for many networks and derives additional hypotheses on how network characteristics affect possibilities for trust (see our related chapter on social networks for a more detailed discussion on game-theoretic models on networks).

Two cautionary remarks on simulation models are in order (see also the chapters by Saam and Opp in this Handbook). First, it is important to realize that data produced using simulations can never replace empirical tests of hypotheses. Analyses on data from simulations can, at best, confirm or strengthen the argumentation that lead to specific hypotheses and are, therefore, an additional method to derive hypotheses from formal models. Data about actual human behavior remain necessary to provide empirical tests of these hypotheses. Second, a danger that easily enters when making models more complex using simulation methods is that too many parameters are added simultaneously and that the models become too complex. A result of this can be that the parameter space cannot be studied systematically enough to ensure that hypotheses about relations between variables are true throughout the parameter space. The advantage of most analytic models is that general theorems are derived that include all relevant variables in the model. We therefore propose that computer simulation should as much as possible be combined with analytical methods, and be aimed at extending analytical results rather than at "simulating reality".

6 Conclusions and suggestions for further reading

In line with the topic of this Handbook, this chapter focused on formal models of social dilemmas and cooperation, with an emphasis on game-theoretic models and also on simulation studies. We have shown how such models can be used to make the notion of a social dilemma precise, to distinguish between important kinds of social dilemmas, and to derive conditions for cooperation in social dilemmas. Specifically, we discussed conditions for cooperation of actors who are incentive-guided and goal-directed, including game-theoretic equilibrium behavior of rational actors. We have emphasized conditions for cooperation of actors who are not only rational but also selfish and have indicated how the selfishness assumption can be replaced by assumptions on social preferences.

The emphasis on game-theoretic models is not accidental. Interdependence between actors is a core feature of a social dilemma: an actor's outcome in a social dilemma is affected not only by his own behavior but also by the behavior of others. Game theory models interdependent situations, providing concepts, assumptions, and theorems that allow to specify how rational actors behave in such situations. The theory assumes that actors behave as if they try to realize their preferences, taking their interdependencies as well as rational behavior of the other actors into account (e.g., Harsanyi 1977). Thus, arguably, game theory is a tailor made tool for sociology, since interdependencies between actors and actors taking their interdependencies into account are likewise the core of Weber's (1947: 88, emphasis added) famous definition of social action: "Sociology [...] is a science which attempts the interpretive understanding of social action in order thereby to arrive at a causal explanation of its course and effects [...] Action is social in so far as [...] *it takes account of the behaviour of others and is thereby oriented in its course.*" Other reviews of research on social dilemmas and cooperation emphasize social psychological theory and other approaches that differ from rational choice assumptions (e.g., Kollock 1998).

Our focus on formal models implied that we indicated testable implications but largely neglected results of empirical research. Buskens & Raub (2013) is an overview that focuses systematically on empirical tests of predictions that follow from game-theoretic models, with an emphasis on using complementary research designs such as experiments, quantitative and qualitative field studies, and quasi-experimental designs such as vignette studies for multiple tests of the same hypotheses (see Levitt & List 2007; Falk & Heckman 2009; and Gächter & Thöni 2011 for further discussion of this issue). Camerer (2003) provides a thorough review of experimental research on social dilemma games and other games from the florishing and rapidly developing field of behavioral game theory. Wittek et al. (2013) is a handbook on empirical applications of rational choice theory in general with various chapters that shed light on social dilemmas and cooperation, too.

References

Arrow, K. (1974) *The Limits of Organization.* New York: Norton.

Axelrod, R. (1984) *The Evolution of Cooperation.* New York: Basic Books.

Balliet, D., L. B. Mulder and P. A. M. van Lange (2011) "Reward, Punishment, and Cooperation: A Meta-Analysis." *Psychological Bulletin* 137: 594–615.

Bendor, J. and D. Mookherjee (1987) "Institutional Structure and the Logic of Ongoing Collective Action." *American Political Science Review* 81: 129–154.

Berg, J., J. Dickhaut and K. McCabe (1995) "Trust, Reciprocity, and Social History." *Games and Economic Behavior* 10: 122–142.

Binmore, K. (1998) *Game Theory and the Social Contract. Volume 2: Just Playing.* Cambridge, MA: MIT Press.

Blau, P. M. ([1964] 1996) *Exchange and Power in Social Life.* New Brunswick, NJ: Transaction Publishers.

Bolton, G. E. and A. Ockenfels (2000) "ERC: A Theory of Equity, Reciprocity and Competition." *American Economic Review* 90: 166–193.

Buskens, V. (2002) *Social Networks and Trust.* Boston, MA: Kluwer.

Buskens, V. and W. Raub (2002) "Embedded Trust: Control and Learning." *Advances in Group Processes* 19: 167–202.

Buskens, V. and W. Raub (2013) "Rational Choice Research on Social Dilemmas: Embeddedness Effects on Trust." Pp. 113–150 in: R. Wittek, T. A. B. Snijders and V. Nee (Ed.) *Handbook of Rational Choice Social Research.* Stanford, CA: Stanford University Press.

Camerer, C. F. (2003) *Behavioral Game Theory. Experiments in Strategic Interaction.* New York: Russell Sage.

Camerer, C. F. and K. Weigelt (1988) "Experimental Tests of a Sequential Equilibrium Reputation Model." *Econometrica* 56: 1–36.

Coleman, J. S. (1964) "Collective Decisions." *Sociological Inquiry* 34: 166–181.

Coleman, J. S. (1987) "Psychological Structure and Social Structure in Economic Models." Pp. 181–185 in: R. M. Hogarth and M. W. Reder (Ed.) *Rational Choice. The Contrast between Economics and Psychology.* Chicago, IL: University of Chicago Press.

Coleman, J. S. (1990) *Foundations of Social Theory.* Cambridge, MA: Belknap Press of Harvard University Press.

Darley, J. M and B. Latané (1968) "Bystander Intervention in Emergencies: Diffusion of Responsibility." *Journal of Personality and Social Psychology* 8: 377–383.

Dasgupta, P. (1988) "Trust as a Commodity." Pp. 49–72 in: D. Gambetta (Ed.) *Trust: Making and Breaking Cooperative Relations.* Oxford: Blackwell.

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

Diekmann, A. (2004) "The Power of Reciprocity." *Journal of Conflict Resolution* 48: 487–505.

Diekmann, A. and S. Lindenberg (2001) "Cooperation: Sociological Aspects." Pp. 2751–2756 in: N. J. Smelser and P. B. Baltes (Ed.) *International Encyclopedia of the Social & Behavioral Sciences. Volume 4.* Amsterdam: Elsevier.

Durkheim, E. ([1893] 1973) *De la Division du Travail Social.* Paris: PUF.

Epstein, J. M. (2006) *Generative Social Science: Studies in Agent-Based Computational Modeling.* Princeton, NJ: Princeton University Press.

Falk, A. and J. J. Heckman (2009) "Lab Experiments Are a Major Source of Knowledge in the Social Sciences." *Science* 326: 535–538.

Fehr, E. and S. Gächter (2000) "Cooperation and Punishment in Public Goods Experiments." *American Economic Review* 90: 980–994.

Fehr, E. and H. Gintis (2007) "Human Motivation and Social Cooperation: Experimental and Analytical Foundations." *Annual Review of Sociology* 33: 43–64.

Fehr, E. and K. M. Schmidt (1999) "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics* 114: 817–868.

Fehr, E. and K. M. Schmidt (2006) "The Economics of Fairness, Reciprocity and Altruism – Experimental Evidence and New Theories." Pp. 615–691 in: S.-C. Kolm and J. M. Ythier (Ed.) *Handbook of the Economics of Giving, Altruism and Reciprocity.* Amsterdam: Elsevier.

Friedman, J. W. (1971) "A Non-Cooperative Equilibrium for Supergames." *Review of Economic Studies* 38: 1–12.

Friedman, J. W. (1990) *Game Theory with Applications to Economics.* New York: Oxford University Press.

Gächter, S. and C. Thöni (2011) "Micromotives, Microstructure, and Macrobehavior." *Journal of Mathematical Sociology* 35: 26–65.

Gambetta, D. (1993) *The Sicilian Mafia. The Business of Private Protection.* Cambridge, MA: Harvard University Press.

Gilbert, N. (2010) *Computational Social Science.* London: Sage.

Gouldner, A. W. (1960) "The Norm of Reciprocity." *American Sociological Review* 25: 161–178.

Granovetter, M. S. (1985) "Economic Action and Social Structure: The Problem of Embeddedness." *American Journal of Sociology* 91: 481–510.

Green, E. J. and R. H. Porter (1984) "Noncooperative Collusion under Imperfect Price Information." *Econometrica* 52: 87–100.

Greif, A. (2006) *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade.* Cambridge: Cambridge University Press.

Güth, W., R. Schmittberger and B. Schwarze (1982) "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior and Organization* 3: 367–388.

Hardin, G. (1968) "The Tragedy of the Commons." *Science* 162: 1243–1248.

Hardin, R. (1982) "Exchange Theory on Strategic Bases." *Social Science Information* 21: 251–272.

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

Harsanyi, J. C. and R. Selten (1988) *A General Theory of Equilibrium Selection in Games.* Cambridge, MA: MIT Press.

Hegselmann, R. (1994) "Solidarität in einer egoistischen Welt – eine Simulation." Pp. 349–390 in: J. Nida-Rümelin (Ed.) *Praktische Rationalität – Grundlagen und ethische Anwendungen des Rational-Choice Paradigmas.* Berlin: De Gruyter.

Hobbes, T. ([1651] 1991) *Leviathan.* Cambridge: Cambridge University Press.

Kahneman, D., J. L. Knetsch and R. Thaler (1986) "Fairness as a Constraint on Profit Seeking: Entitlements in the Market." *American Economic Review* 76: 728–741.

Klein, D. B. (1997) *Reputation: Studies in the Voluntary Elicitation of Good Conduct.* Ann Arbor, MI: University of Michigan Press.

Kollock, P. (1998) "Social Dilemmas: The Anatomy of Cooperation." *Annual Review of Sociology* 24: 183–214.

Kreps, D. M. (1990a) "Corporate Culture and Economic Theory." Pp. 90–143 in: J. Alt and K. Shepsle (Ed.) *Perspectives on Positive Political Economy.* Cambridge: Cambridge University Press.

Kreps, D. M. (1990b) *Game Theory and Economic Modelling.* Oxford: Clarendon Press.

Ledyard, J. O. (1995) "Public Goods: A Survey of Experimental Research." Pp. 111–194 in: J. H. Kagel and A. E. Roth (Ed.) *The Handbook of Experimental Economics.* Princeton, NJ: Princeton University Press.

Levitt, S. D. and J. A. List (2007) "What Do Laboratory Experiments Measuring Social Preference Reveal About the Real World?" *Journal of Economic Perspectives* 21: 153–174.

Lindenberg, S. (1992) "The Method of Decreasing Abstraction." Pp. 3–20 in: J. S. Coleman and T. J. Fararo (Ed.) *Rational Choice Theory. Advocacy and Critique.* Newbury Park, CA: Sage.

Lindenberg, S. (2001) "Social Rationality versus Rational Egoism." Pp. 635–668 in: J. Turner (Ed.) *Handbook of Sociological Theory.* New York: Kluwer Academic/Plenum.

Macaulay, S. (1963) "Non-Contractual Relations in Business." *American Sociological Review* 28: 55–66.

Macy, M. W. (1991) "Learning to Cooperate: Stochastic and Tacit Collusion in Social Exchange." *American Journal of Sociology* 97: 808–843.

Macy, M. W. and A. Flache (2002) "Learning Dynamics in Social Dilemmas." *Proceedings of the National Academy of Sciences USA* 99: 7229–7236.

Milgrom, P., D. C. North and B. R. Weingast (1990) "The Role of Institutions in the Revival of Trade: The Law Merchants." *Economics and Politics* 2: 1–23.

Nowak, M. A. (2006) "Five Rules for the Evolution of Cooperation." *Science* 314: 1560–1563.

Olson, M. (1965) *The Logic of Collective Action.* Cambridge, MA: Harvard University Press.

Ortmann, A., J. Fitzgerald and C. Boeing (2000) "Trust, Reciprocity, and Social History: A Re-Examination." *Experimental Economics* 3: 81–100.

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

Parsons, T. (1937) *The Structure of Social Action.* New York: Free Press.

Rapoport, A. (1974) "Prisoner's Dilemma – Recollections and Observations." Pp. 18–34 in: A. Rapoport (Ed.) *Game Theory as a Theory of Conflict Resolution.* Dordrecht: Reidel.

Rasmusen, E. (2007) *Games and Information: An Introduction to Game Theory.* (4th edition) Oxford: Blackwell.

Raub, W. (1988) "Problematic Social Situations and the 'Large-Number Dilemma'." *Journal of Mathematical Sociology* 13: 311–357.

Raub, W. (2004) "Hostage Posting as a Mechanism of Trust: Binding, Compensation, and Signaling." *Rationality and Society* 16: 319–366.

Raub, W., V. Buskens and M. A. L. M. van Assen (2011) "Micro-Macro Links and Microfoundations in Sociology." *Journal of Mathematical Sociology* 35: 1–25.

Raub, W. and T. Voss (1986) "Die Sozialstruktur der Kooperation rationaler Egoisten. Zur 'utilitaristischen' Erklärung sozialer Ordnung." *Zeitschrift für Soziologie* 15: 309–323.

Van de Rijt, A. and M. W. Macy (2009) "The Problem of Social Order: Egoism or Autonomy?" *Advances in Group Processes* 26: 25–51.

Rosenthal, R. W. (1981) "Games of Perfect Information, Predatory Pricing and the Chain-Store Paradox." *Journal of Economic Theory* 25: 92–100.

Sally, D. (1995) "Conversation and Cooperation in Social Dilemmas: A Meta-Analysis of Experiments from 1958 to 1992." *Rationality and Society* 7: 58–92.

Schelling, T. C. (1960) *The Strategy of Conflict.* London: Oxford University Press.

Schelling, T. C. (1978) *Micromotives and Macrobehavior.* New York: Norton.

Schramm, M. and M. Taube (2003) "Evolution and Institutional Foundation of the Hawala Financial System." *International Review of Financial Analysis* 12: 405–20.

Sen, A. (1974) "Choice, Orderings and Morality." *Choice, Welfare and Measurement.* Oxford: Blackwell.

Snijders, C. (1996) *Trust and Commitments.* Amsterdam: Thesis Publishers.

Snijders, C. and G. Keren (1999) "Determinants of Trust." Pp. 355–385 in: D. V. Budescu, I. Erev and R. Zwick (Ed.) *Games and Human Behavior.* Mahwah, NJ: Lawrence Erlbaum.

Snijders, C. and G. Keren (2001) "Do You Trust? Whom Do You Trust? When Do You Trust?" *Advances in Group Processes* 18: 129–160.

Squazzoni, F. (2012) *Agent-Based Computational Sociology.* Chichester: Wiley.

Stigler, G. J. (1964) "A Theory of Oligopoly." *Journal of Political Economy* 72: 44–61.

Taylor, M. ([1976] 1987) *The Possibility of Cooperation.* Cambridge: Cambridge University Press. Rev. ed. of *Anarchy and Cooperation.* London: Wiley.

Ullmann-Margalit, E. (1977) *The Emergence of Norms.* Oxford: Clarendon.

Vogt, S. and J. Weesie (2004) "Social Support among Heterogeneous Partners." *Analyse & Kritik* 26: 398–422.

Voss, T. (1982) "Rational Actors and Social Institutions: The Case of the Organic Emergence of Norms." Pp. 76–100 in: W. Raub (Ed.) *Theoretical Models and Empirical Analyses. Contributions to the Explanation of Individual Actions and Collective Phenomena.* Utrecht: ESP.

Walster, E., W. G. Walster and E. Berscheid (1978) *Equity. Theory and Research.* Boston, MA: Allyn and Bacon.

Weber, M. ([1921] 1976) *Wirtschaft und Gesellschaft.* Tübingen: Mohr.

Weber, M. (1947) *The Theory of Social and Economic Organization.* New York: Free Press.

Weesie, J. (1988) *Mathematical Models for Competition, Cooperation, and Social Networks.* Utrecht: Diss. Utrecht University.

Weesie, J. and W. Raub (1996) "Private Ordering: A Comparative Institutional Analysis of Hostage Games." *Journal of Mathematical Sociology* 21: 201–240.

Weesie, J., V. Buskens and W. Raub (1998) “The Management of Trust Relations via Institutional and Structural Embeddedness.” Pp. 113–138 in: P. Doreian and T. Fararo (Ed.) *The Problem of Solidarity: Theories and Models.* Amsterdam: Gordon and Breach.

Weibull, J. (1995) *Evolutionary Game Theory.* Cambridge, MA: MIT Press.

Williamson, O. E. (1985) *The Economic Institutions of Capitalism.* New York: Free Press.

Wittek, R., T. A. B. Snijders and V. Nee (Ed.) (2013) *Handbook of Rational Choice Social Research.* Stanford, CA: Stanford University Press.