Ben Jann and Wojtek Przepiorka

Introduction

The question of how cooperation and social order can evolve from a Hobbesian state of nature of a "war of all against all" (Hobbes [1668] 1982) has always been at the core of social scientific inquiry (e.g., Axelrod 1984; Bowles and Gintis 2011; Durkheim [1893] 1997; Ellickson 1991; Ostrom 1990; Sennett 2012). Various approaches exist for addressing this question, but the theoretical lens through which we view the phenomena presented in this book is methodological individualism (Weber [1922] 2013; Coleman 1990). Methodological individualism reminds us that cooperation and social order are macro-sociological phenomena that can be, and need to be, explained as a result of the goal-oriented behaviors of actors. A key insight from this point of view is that *social dilemmas* constitute a pivotal analytical paradigm that can be used by social scientists to investigate the origins of conflict, competition, and cooperation among humans. Social dilemmas are therefore well suited as the micro-foundational building blocks for a stringent theoretical explanation of cooperation and social order (Kollock 1998).

Social dilemmas are "situations of strategic interdependence in which the decisions of individually rational actors lead to an inferior outcome for all or some parties than the decisions of 'collectively rational' actors. Collective rationality means that actors, had they an opportunity to communicate and agree on a binding contract, should agree on a combination of actions leading to a welfare-enhancing outcome" (Diekmann and Przepiorka 2016:1311). Social dilemmas, therefore, are at the core of the problem of social cooperation. Their analysis identifies when and why humans may struggle to achieve a collectively rational solution, what societal benefits they forgo in their struggle, and what mechanisms and measures may help overcome the struggle.

Social dilemmas are usually described in game theoretic terms. Apart from the well-known Prisoner's Dilemma, there are many examples of social dilemmas (McAdams 2009; Raub, Buskens, and Corten 2015). Social dilemmas are mostly studied in small groups, such as dyadic interactions with trust at stake, or voluntary contribution situations in which a public good is produced only if enough group members contribute their resources. However, social dilemmas also occur on a large scale. The common pool resource dilemma is often used to describe the clash between individual and collective rationality in environmental issues such as fishery, land use, and traffic (Ostrom 1990). Coordination problems, in which actors must agree on one of many welfare-enhancing outcomes, also fall under the definition of social dilemmas.

The gap between the individual and collective rationality inherent in social dilemmas (Rapoport 1974) creates a demand for the regulation of actors' behavior (Coleman 1990; Voss 2001). Such regulations can be formal and manifest themselves in terms of legal codes and other institutions providing selective incentives. Regulations can also,

https://doi.org/10.1515/9783110472974-001

however, be informal, and emerge as social norms that prescribe or proscribe certain behavior that is enforced by positive and negative peer-sanctions (Hechter and Opp 2001).

In general, whether formal or informal, such regulations can be understood as *institutions*, defined as "humanly devised constraints that shape human interaction" (North 1990:3). A specific type is social norms, defined as rules guiding social behavior, the deviation from (or adherence to) which is negatively (or positively) sanctioned (e.g., Bicchieri 2006). Both social norms and other institutions are subject to change, due to technological innovations, policies and external shocks such as economic, political and environmental crises. These changes, in turn, affect the ways in which we interact on a small and on a large scale, and how we cooperate. Thus, while the notion of social dilemmas makes problems of cooperation comprehensible, the analysis of institutions is key to understanding the *evolution of cooperation*.

Against this backdrop, the chapters compiled in this book give an overview of state-of-the-art research on social dilemmas, institutions, and the evolution of cooperation. The book covers (1) theoretical analyses of social dilemmas such as trust, public goods, common-pool resource and coordination dilemmas; (2) the role of formal and informal organizations, social norms, and institutions in shaping how individuals interact to overcome social dilemmas; and (3) empirical studies conducted in the laboratory and the field, as well as agent-based simulations that investigate how cooperation evolves in human groups. The book is divided into seven parts, and comprises 26 chapters written by distinguished scholars and experts in the field.

Part I: Introduction

After this introductory chapter, the book begins with an article by Raub and Voss (Chapter 2), who remind us of the theoretical roots of European rational-choice sociology. While James Coleman can safely be called the father of sociological rational choice theory, some of his ideas, and in particular the famous Coleman-boat, turn out to have originated in Europe. Raub and Voss provide an overview of the European antecedents of the Coleman-boat and thereby place the contributions in this book against the context of the history of thought. At the same time, by discussing various micro-macro models, Raub and Voss introduce the basic ideas of methodological individualism.

Part II: Institutions

The second part starts with a description of one of the most peculiar institutions in human social history – the Kula Ring. The Kula Ring was a system of economic and ceremonial exchanges among islander tribes in the Western Pacific, and was first described by the ethnographer Bronisław Malinowski at the beginning of the last century. It is believed that the ceremonial gift exchanges that took place alongside economic exchanges served the function of maintaining peace among the tribal societies.

How the Kula Ring came into existence, however, is a yet unsolved puzzle. Ziegler (Chapter 3) devises a simulation model that aims at identifying the conditions under which the Kula Ring may have evolved.

Perhaps less peculiar, but at least as important in terms of their origins as the Kula Ring, are criminal justice institutions. Starting from the observation that there is substantial cross-cultural variability in what is considered to be a crime, Eisner, Murray, Ribeaud, Averdijk, and van Gelder (Chapter 4) challenge the widespread view that the evolved human psychology has had a major bearing on the prevalence of criminal justice institutions. The authors emphasize procedural fairness and legitimacy as inherent properties of criminal justice institutions with a positive effect on maintaining social order. These properties distinguish criminal justice institutions from mere centralized punishment systems.

In a similar vein, Lindenberg (Chapter 5) argues that institutions do not regulate actors' behavior via selective incentives alone, but also have a bearing on actors' perceptions of the situation at hand, their beliefs about other actors, and their own motivations. The author employs goal-framing theory to demonstrate how institutions regulate actors' behavior via their influence on these actors' overarching goals. He argues that the functioning of institutions crucially depends on how they influence an actor's three overarching normative, gain and hedonic goals. To function well, institutions must make actors' normative goals salient and activate internalized norms to prompt actors to act according to the legitimate rules these institutions define.

In the last chapter in Part II, Mueller (Chapter 6) gives several examples of large-scale social dilemmas related to vaccinations against infectious agents, and shows how the use of social dilemma theory can help to devise institutions that solve the dilemmas inherent in vaccination decisions.

Part III: Social Norms

What one ought to do often clashes with what one would like to do. In Chapter 7, Opp challenges the common view that the effects of one's inclination to follow social norms and one's own interests on behavior are additive. He suggests instead that one's inclination to follow social norms decreases with one's inclination to follow one's own interests (and vice versa). He uses survey data on protest participation during the East German Revolution in 1989 to test his proposition about the interaction effect of social norms and own interests.

Own interests, however, may also instigate collective action and normative change. In his contribution, Preisendörfer (Chapter 8) addresses the question of whether adverse environmental conditions trigger affected actors' activism, aiming at improving their environmental conditions and changing social norms through policy interventions. To test his proposition, the author uses the Swiss Environmental Survey, which combines answers to the survey questions with objectively measured environmental conditions in the neighborhood of the respondents.

Social norms change as a result of institutional change. In academia, large-scale cooperative projects involving many researchers are increasingly promoted. At the same time, individual researchers compete for scarce, long-term academic positions and are evaluated based on their individual performance. Cooperating with others on a common project or working alone is therefore a dilemma with which young researchers are increasingly confronted. Gross, Jungbauer–Gans, and Nisic (Chapter 9) investigate whether the norm to cooperate with others has increased over time, and whether researchers who cooperate with others are more successful in securing long-term academic positions.

Another clash between social norms and own interests in academia manifests itself every now and then in prominent cases of scientific misconduct. For social norms regarding scientific conduct to be established and maintained, norm violations must be negatively sanctioned with a certain probability; detecting and documenting scientific misconduct is a necessary precondition for the application of sanctions. In Chapter 10, Auspurg and Hinz devise and empirically evaluate new methods of fraud detection in scientific publications.

Reciprocity is arguably the strongest social norm gluing a society together. Liebe and Tutic (Chapter 11) conduct two quasi-experiments to explore the interplay of reciprocity and social status. They use the sequential dictator game to measure reciprocity and manipulate social status by allowing participants from different schools to interact with each other.

Part IV: Peer-Sanctioning

Peer-sanctioning can be effective in establishing norm compliance, if a significant proportion of actors are willing to sanction their deviant peers at a cost to themselves. However, the effectiveness of peer-sanctioning may not only depend on the number of potential sanctioners, but also on the conflict potential inherent in actors' divergent normative expectations. In Chapter 12, Rauhut and Winter distinguish four types of normative conflict that can arise from actors' diverging expectations regarding the type of a social norm, and the extent to which others should adhere to that norm. Additionally, the more the actors to whom the social norm is directed are distinct from the actors who benefit from the social norm, the less effective peer-sanctioning will be in establishing norm compliance.

In line with the idea of normative conflict, Jann and Coutts (Chapter 13) provide empirical evidence of the extent of negative peer-sanctioning in an inventive field experiment. They show how larger differences in actors' social status can lead to more aggressive peer-sanctions of traffic norm violations. In their experiment, a confederate seemingly unintentionally blocks a road with his or her car, and the time until the blocked driver honks is measured. Actors' social status is measured by the type of car.

A factor that has been shown to hamper the effectiveness of peer-sanctioning to promote norm compliance is negative counter-sanctioning (i.e., retaliation for pun-

ished deviations). In Chapter 14 Flache, Bakker, Mäs and Dijkstra suggest that this may also be true for positive counter-sanctioning (i.e., reciprocation of rewarded adherence). The authors stage a laboratory experiment in which they test this conjecture by varying whether subjects can reward or punish their peers, and whether these sanctions occur anonymously or are attached to subjects' "identities" for the duration of the experiment.

Winter and Franzen (Chapter 15) take up the following important but often neglected question concerning peer-sanctioning: who is going to sanction the norm breaker? A group of actors who experience a norm violation by another actor may face a coordination dilemma in which only one actor is required to sanction the norm breaker. The authors show that this coordination dilemma can lead to the diffusion of responsibility such that the likelihood of a sanction decreases with the number of actors. They test this hypothesis by means of a laboratory experiment with the multi responder ultimatum game.

In the last decade or so, research has shown how centralized sanctioning institutions could have evolved to substitute the costly and therefore often inefficient peer-sanctioning mechanism. In these studies, subjects can choose whether they want to be in an environment with or without the possibility of peer-sanctioning. Environments employing a peer-sanctioning institution turn out to be more successful and therefore attract more actors over time. In Chapter 16, van Miltenburg, Buskens, and Raub conduct a laboratory experiment to investigate whether this result also holds if the cooperation of actors is observed by their group members with some noise; with noise, defectors are observed as cooperators (and vice versa) with a small probability.

Part V: Trust and Trustworthiness

While social norms have been called the "cement" of society, trust has been called society's "lubricant". If actors know they can trust each other, social dilemmas can be overcome more effectively because fewer transaction costs accrue from the regulation of actors' behavior. Under what conditions, however, can and do actors trust each other? Taking a psychological perspective, Oswald and Ulshöfer (Chapter 17) distinguish between trust and distrust as two states of mind that are activated depending on the risks involved in a social dilemma. In situations with small stakes, trust is the default state of mind, whereas distrust is the default in situations with larger stakes. The authors argue that actors are more skeptical about their interaction partners in the state of distrust. This in turn can have a positive effect, as actors are less gullible and more alert to signs and signals of trustworthiness in their interaction partners.

Conversely, consistent with a rational choice perspective, Przepiorka and Berger (Chapter 18) start with the assumption that actors are in a state of mind in which they process information about their interaction partners' trustworthiness in an accurate way, irrespective of whether stakes are low or high. After giving a precise definition of the trust dilemma, Przepiorka and Berger outline the potential of signaling theory

to explain trust and trustworthiness in social exchange. They argue that the conceptual distinction between signals and signs, and between the production and display of signals and signs, can make signaling theory more broadly applicable in the social sciences in general and in sociological scholarship in particular.

Vieth and Weesie (Chapter 19) make the point more explicit that not only stakes determine whether someone is trustworthy or trustful, but also one's interaction partners' discernable intentions. That is, the mere act of trusting someone and promises of trustworthiness can induce actors to be respectively more trustworthy or more trustful. The authors conduct a laboratory experiment to test these conjectures, employing a nested game design, which allows disentangling subjects' motives and intentions in the trust dilemma.

In the last chapter of Part V, Snijders, Bober, and Matzat (Chapter 20) highlight once more why it is so important to study the causes of trust and trustworthiness. More and more social interactions are taking place online. In peer-to-peer online markets such as eBay, anonymous traders exchange goods and services across large geographic distances. The functioning of these online market platforms crucially depends on electronic reputation systems, which allow traders to rate each other after finished transactions and, in this way, to produce valuable information about the trustworthiness of potential exchange partners. The authors conduct an online choice experiment in which they present subjects with eBay-like offers of digital cameras to address two important but understudied questions: how much do positive and negative text messages, as compared to positive and negative star-ratings, affect potential buyers' trust in online sellers, and how can sellers rebuild their trustworthiness in their responses to negative text messages?

Part VI: Game Theory

The three chapters in this part, although partly inspired by empirical evidence that conflicts with theoretical predictions, are purely game theoretical. Nax, Murphy, and Helbing (Chapter 21) make a case for the reconsideration of simple learning models to explain actors' behavior in social dilemmas. In their analysis, they focus on a set of public goods dilemmas and the meritocratic matching mechanism. Meritocratic matching is a mechanism by which cooperators tend to be grouped with cooperators and defectors tend to be grouped with defectors, thereby making cooperation among self-regarding actors possible under certain conditions. The authors argue that simple learning models explain actors' behavior in these types of games better than preference-based models.

In a step-level public goods dilemma, the voluntary contribution of m actors is required to produce the public good for the entire group of $n \geq m$ actors. It is usually assumed that m is common knowledge. In his analysis, Bolle (Chapter 22) assumes that m is only known to a requestor (e.g., editors asking authors for their contributions to a collective volume) who is not part of the group, can communicate this information

to the other actors, and benefits if the public good is produced. Bolle outlines the conditions under which the requestor has an incentive strategically to misrepresent the required number of volunteers.

In the last chapter of Part VI, Gautschi (Chapter 23) analyses the trust dilemma with hostage posting. "Hostages" are costly commitment devices that can be used by actors to overcome the trust dilemma. The author observes that in experiments with hostage trust games, not posting a hostage results in less cooperation than in a trust dilemma, in which hostage positing is not an option. Given the equivalence of the trust dilemma without hostage posting, and the trust dilemma with hostage posting in which no hostage is placed, this empirical finding needs an explanation.

Part VII: Experimental Methods

Framing-effects are empirical effects of the "name of the game". For example, actors behave differently in an otherwise identical social dilemma, which in one case is called the community game and in another is called the Wall Street game. Esser (Chapter 24) reviews the different ways in which experimental social scientists with a proclivity for rational choice theory have reacted to framing-effects. He thereby introduces the so-called model of frame selection, which offers a way to integrate framing-effects into a broader notion of rational choice theory.

Actors with social preferences are more inclined to cooperate in social dilemmas, but not all actors have social preferences and actors' preferences are not directly observable. However, institutions that take actors' preferences into account can be more effective in providing solutions to social dilemmas. The design of such institutions thus requires that actors' preferences can be measured. In Chapter 25, Höglinger and Wehrli evaluate the reliability and validity of a recently devised measure of social preferences, the SVO slider measure, by means of an online experiment.

Laboratory experiments with university students as participants have been the main approach to empirical research on social dilemmas. Berger and Baumeister (Chapter 26) address the question of how far our knowledge about behavior in social dilemmas could be influenced, and even biased, by the way laboratory experiments are conducted. In particular, the authors look at the effect of subjects' repeated participation in similar laboratory experiments on these subjects' behavior in social dilemmas.

Bibliography

[1] Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Basic Books.
[2] Bicchieri, Christina. 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press.

[3] Bowles, Samuel, and Herbert Gintis. 2011. *A Cooperative Species: Human Reciprocity and its Evolution*. Princeton: Princeton University Press.
[4] Coleman, James S. 1990. *Foundations of Social Theory*. Cambridge, MA: The Belknap Press of Harvard University Press.
[5] Diekmann, Andreas, and Wojtek Przepiorka. 2016. "Take One for the Team!" Individual heterogeneity and the emergence of latent norms in a volunteer's dilemma." *Social Forces* 93(3):1309–1333.
[6] Durkheim, Emile. [1893] 1997. *The Division of Labor in Society*. New York: Free Press.
[7] Ellickson, Robert C. 1991. *Order without Law: How Neighbors Settle Disputes*. Cambridge: Harvard University Press.
[8] Hechter, Michael, and Karl-Dieter Opp, eds. 2001. *Social Norms*. New York: Russell Sage Foundation.
[9] Hobbes, Thomas. [1668] 1982. *Leviathan*. London: Penguin Classics.
[10] Kollock, Peter. 1998. "Social Dilemmas: The Anatomy of Cooperation." *Annual Review of Sociology* 24:183–214.
[11] McAdams, Richard H. 2009. "Beyond the Prisoners' Dilemma: Coordination, Game Theory, and Law." *Southern California Law Review* 82(2):209–258.
[12] North, Douglass C. 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.
[13] Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.
[14] Rapoport, Anatol. 1974. "Prisoner's Dilemma – Recollections and Observations." Pp. 17–34 in *Game Theory as a Theory of Conflict Resolution*, edited by A. Rapoport. Dordrecht: Reidel.
[15] Raub, Werner, Vincent Buskens, and Rense Corten. 2015. "Social Dilemmas and Cooperation." Pp. 597–626 in *Handbuch Modellbildung und Simulation in den Sozialwissenschaften*, edited by N. Braun, and N. J. Saam. Wiesbaden: Springer VS.
[16] Sennett, Richard. 2012. *Together: The Rituals, Pleasures and Politics of Cooperation*. New Haven: Yale University Press.
[17] Voss, Thomas. 2001. "Game-Theoretical Perspectives on the Emergence of Social Norms." Pp. 105–136 in *Social Norms*, edited by M. Hechter, and K.-D. Opp. New York: Russell Sage Foundation.
[18] Weber, Max. [1922] 2013. *Economy and Society*. Berkeley: University of California Press.