How we cooperate, John E. Roemer, Yale University Press 2019, 248 pages

As I write this, we are living in strange times, through the COVID-19 pandemic and lockdown. The newspapers are filled with headlines about R numbers, excess mortality, and saving the economy versus saving lives. But some things about the situation have a comforting academic familiarity: we are living in a giant public goods game. One person leaving the house is of no danger to anyone, but if everyone goes outside there is a crowd and the virus may spread. Or perhaps it is less comforting to think of oneself as lab rat, in a permanent experiment aiming to achieve the optimal amount of social contact. Nevertheless, people have been astoundingly compliant with social isolation—surprisingly compliant according to standard economics, although less surprising to behavioural economists and researchers in other disciplines.

The messaging to encourage compliance has been diverse. One prominent message is to 'protect yourself and your family'. Alternatively, the UK Secretary of State for Health and Social Care has been exhorting people to 'do their duty'. The first of these messages expands self-interest to include pro-social motivations, the second appeals to moral principles. Roemer's theory in *How we cooperate* falls firmly into the second camp. He objects to the approach taken in behavioural economics, of modifying classical self-regarding preferences to incorporate other-regarding motivations, such as altruism or fairness. He argues that, instead of introducing non-standard preferences, we should introduce non-standard optimization, inspired by Kant's categorical imperative, to act only in ways that one can wish would become a universal law. Roemer introduces a new equilibrium concept and shows how it can supply general microfoundations for cooperative behaviour, in both simple games and whole economies.

Roemer motivates his approach by lamenting the deficiencies of behavioural economics and it is worth listing those here, since they also provide a yardstick for judging the success of his proposal. Roemer complains that behavioural economics provides clear solutions in simple games with a restricted set of options (such as games with 2x2 matrices), or games where the cooperative solution is obvious. However, when the game is complex and the cooperative profile isn't obvious, then he contends that people will struggle to apply the solution concepts. He also has a secondary concern that behavioural economics is unfalsifiable. Roughly the idea seems to be that, if anything can be an argument in the preference function, especially if we allow parameterization at the level of individual players or individual games, then everything can be explained retrospectively. Fitting the preferences to the actions is the equivalent of data mining in econometrics.

Roemer also has the intuition that theories of non-self-regarding preferences do not capture our motivation for cooperating. He proposes that people cooperate because they want to do the right thing, where 'do the right thing' is defined in large part by what others do: cooperation is a *quasi-moral norm* whose content is 'cooperate if *and only if* most others do' (Elster, 2017, p.728). Therefore, Roemer's agents want to do the right thing, provided that they expect that others will act in a like manner. Further, if we are all situated similarly, then the right action will be the same for everyone. This leads Roemer to define a simple Kantian equilibrium (SKE) as the strategy that *each* would prefer *all* to play in a symmetric game (or any non-symmetric game that can be re-conceptualised as symmetric, see below).

This is equivalent to changing the question that the agent asks. A standard Nash optimizer asks 'Given the strategy chosen by my opponent, what is the best strategy for me?'. Behavioural economics modifies how an agent decides what is 'best for me', expanding self-interest to take account of other players' outcomes. In contrast, a Kantian optimizer asks, 'What is the strategy I would like both of us to play?'. The Kantian optimizer is still asking about what is best for her; the outcome of the other player is not relevant, she has no other-regarding motivation. What has changed is the way the Kantian optimizes: instead of holding others' actions constant and seeing what happens when she changes her own, the Kantian considers only outcomes where everyone will behave in the same way that she does. Roemer says that, where Nash players treat other players' behaviour as a part of the environment, a Kantian player regards other players as a part of the action.

Kantian optimization can produce an efficient solution in games with externalities. Take the prisoner's dilemma. Both players have to do the same thing, so the off-diagonal strategies are knocked out, and with them the possibility of defecting to gain at the other's expense. Kantian players must choose between 'all cooperate' and 'all defect' and they can reason that 'I prefer that we both play cooperate than that we both play defect' and the SKE is therefore (cooperate, cooperate). This reasoning also leads to SKEs of (stag, stag) in the stag hunt and (dove, dove) in hawk dove. (All three games are shown in matrix form in Fig. 1.)

Roemer also applies SKE to simple asynchronous games, where the players make their moves at different times, such as the dictator game, the ultimatum game and the trust game. These games can be made symmetric if we add a first move, where Nature chooses the order of the players. Then Roemer suggests that a Kantian optimizer asks, 'How would I like each of us to play if each of us could be chosen to be the first player?'. Fairness enters as the realisation that Nature could have chosen either player to be first. In a dictator game, if Nature chooses each player to be the dictator with probability one half, then the SKE has the dictator offering half the prize. Similarly, in the ultimatum game, Roemer argues that a Kantian should think, 'If I were chosen to be the ultimator, and were to propose to keep *x*, this must be the amount I would also like the other person to keep, were she chosen to move first, and hence I must accept any amount from her that is greater than 1 - *x*'. Then at the unique SKE each player gets half the prize. In the SKE of a trust game, the first player sends over her entire endowment and the second player returns at least as much money as he was sent.

Roemer proves the conditions for the existence and efficiency of a SKE in games. A SKE will exist if there is a common diagonal, i.e. if each player has the same ordering over the payoffs on the main diagonal. It will be efficient in strictly monotone games, where the payoff to a strategy is either strictly increasing or strictly decreasing in the number of players who play it; in other words, in games where the strategies are strategic complements.

The limitations of these conditions become clear when we consider another simple game from Game Theory 101, the battle of the sexes. The players want to be together but they have strictly opposing views about where they should go, if they go to the same place. There is no common diagonal in the standard formulation of the game (see Fig. 2a). Player 1 prefers that they go together to the dance and Player 2 prefers that they go to the boxing match. Roemer uses a variant of the game where the players get a strictly positive utility if they go to their preferred event alone, though less than they would get from going to either event together (see Fig. 2b). We can create a common diagonal if we re-label the strategies 'go to the event I prefer' and 'go to the event my partner prefers' (see Fig. 2c). The SKE in pure strategies is that each goes to their preferred event, which is not Pareto efficient because both players prefer going to the same event. However, Roemer shows that the mixed-strategy SKE (where the players assign each strategy a strictly positive probability and then randomize over them according to those probabilities) strictly dominates the mixed-strategy Nash equilibrium, basically because the Kantian optimizers are willing to compromise more and go to the other player's preferred activity with a higher probability. More generally, Roemer shows that the SKE dominates the symmetric Nash equilibria in all cases where they do not coincide.

I will leave aside general concerns about mixed strategy equilibria (e.g. that they are unstable and that real people are not very good at randomising) and focus on the fact that SKE is an equilibrium under a description. When Roemer advocates re-describing the strategies in order to create a common diagonal, I am reminded of David Lewis's discussion in *Convention* of what it is to do the same action (Lewis, 1969). In Lewis's example, two people on the phone get disconnected and would like to resume the call, which requires coordinating so that only one of them makes a call (if both call then they will get engaged signals). Should we name the strategies 'calling back' and 'not calling back', or 'call back if and only if one is the original caller' and 'call back if and only if one is not the original caller'? The move from the first set of descriptions to the second creates a common diagonal. However, Lewis rejects the idea that coordination is doing whatever the others do, on the grounds that 'this is a dangerous thing to say, since it is true of a coordination problem only under suitable descriptions of actions, and sometimes the descriptions that make it true would strike us as contrived' (p.12). He goes on to define coordination in a way that does not depend on the description of the players. Roemer takes the path that Lewis forsook: Kantian optimizers are trying to do the same as what others will do, so a SKE is an equilibrium under some description and the descriptions we choose will matter, inviting the debate over descriptions that Lewis side-stepped.

What is the right description of a game? It seemed quite simple in Roemer's 2x2 battle of the sexes game. However, the problem is going to get harder as the game gets more complicated and there are more strategies available. For instance, imagine that the battle of the sexes game is the stage game of a repeated game. The mixed strategy equilibrium is not going to be efficient compared to a rule that ensures the players coordinate, for instance alternating between events. But now there is a coordination problem that symmetry cannot resolve. Should they alternate events starting with ballet or alternate events starting with boxing? Even if they do solve the coordination problem and both go to the boxing first, why alternate every other event rather than, for instance, going to the boxing twice and then the dance twice ad infinitum? I can begin to think of some arguments that one could construct for alternating every other event, especially if we allow time preferences to enter. However, apart from not solving the coordination problem and therefore not being efficient, it no longer looks like Kantian optimisation gives a simple and obvious solution, which was one of the motivations for abandoning the approach of behavioural economics.

Further, once we introduce repetition, there are an infinity of possible rules and, if any symmetric rule is a candidate for the SKE, then we seem to have the same problem of flexibility in interpretation that Roemer claimed plagues behavioural economics. The idea of 'do whatever I would like everyone to do' seems simple and generalisable, but we need a more precise description of the strategies so that players can all do the same thing, which will differ for different games. Any behaviour can be explained retrospectively by fitting the rule to the game.

Indeed, the choice of rule does matter when Roemer moves onto production economies, where individuals exert effort in order to produce an output that is shared between them according to an allocation rule. Examples of allocation rules are distributing output according to effort and equal division of output amongst agents. When considering how much effort to exert, someone can consider whether to increase effort by 10% in which case the Kantian optimizer would consider whether she would like it if everyone expanded their labour by 10% (*proportional rule*). Or she can consider whether to take a two-hour nap, in which case the Kantian would consider whether she would be happy if everyone took a two-hour nap (*additive rule*). Again, Roemer shows that Kantian optimisation can lead to an efficient amount of effort being expended when there are externalities. However, the catch is that the SKE will only be efficient if the Kantian rule matches the allocation rule. The proportional rule is efficient in an economy that allocates output in proportion to effort, but the additive rule is efficient in an economy that splits the output equally amongst agents. If the allocation rule and the Kantian rule do not match, then the Kantian equilibrium will not be efficient. How the players cash out 'all do the same thing' matters.

Roemer sees this problem and floats three different responses. The first is that anthropology suggests that there is cultural variation, so not all societies will find the efficient rule. The second is that evolution will favour societies that use the efficient rule. The third is that he is not too bothered by this problem because he thinks that the Kantian equilibrium should mainly be viewed as prescriptive, as instructions for how a group that wants to cooperate can design an attractive solution. This last seems a bit disingenuous, given the way that Roemer has motivated his project and that he devotes space to presenting and discussing some survey data he collected on people's motivation for everyday cooperative activities, such as recycling and voting.

(The results are consistent with Kantianism, but also with a number of theories; Elster, 2017, provides a detailed critique.) In the time-honoured tradition of rational choice theory, Kantian optimization waivers between being a descriptive theory and a normative one.

Unlike Lewis (1969), I'm not committed to finding solutions that are description-invariant; I think there are cases where we need to model the agent's description of the situation. Unlike Roemer, I have some optimism that we can test descriptive theories of behaviour in the laboratory. My intuition is that the varied examples on Roemer's list of cooperative behaviours, which run from recycling to charitable giving, may be driven by different psychological mechanisms or even by different motivations in different people. It may be that Roemer's theory is not successful according to his own yardsticks, but since I believe in an ecology of motivations, I am happy to have another plausible theory to add to the contenders.

To finish, I want to compare Kantian optimization to another theory that solves problems of cooperation by changing the question. Theories of team agency allow that groups can be agents and, instead of asking 'What should I do?', a player can ask 'What should we do?' (Sugden, 1993; Bacharach, 2006). A team reasoner considers which combination of actions by members of the team would best promote the team's objective and then performs her part of that combination. Recasting this in Roemer's language, team reasoning changes the question from 'What is the best strategy for me? to 'What is the best set of strategies for us?'. This is the general structure of a family of theories, each of which has a different way of filling in the conditions under which team reasoning occurs, and the relationship between the individual and the team's preferences (Gold, 2012).

Allowing that a group can be an agent has parallels with Roemer's idea that a cooperator regards other individuals as a part of the action. But team reasoning is not limited to symmetric games. Potentially, Kantian optimization could be seen as a type of team reasoning, which occurs when there is common knowledge that others use it and restricts the team reasoner to considering only symmetric strategy profiles, which generates a unanimous ranking of outcomes.

In this review, I have only scratched the surface of the book. I have focused on chapters in Part 1, which also covers topics such as the evolution of Kantian optimization, and have not talked at all about Part 2, which: proves the existence and dynamics of a Kantian equilibrium in production economies; shows how Kantian optimization can solve problems of market economies, including global carbon emissions and efficient provision of publics goods; and ends with a vision of how a market socialist economy could work, with either citizen- or worker-owned firms. I have also focused on the more prose-heavy part of the book, since Part 2 is a succession of models without the sociological and anthropological case studies that are offered as just-so stories and which give some colour to Part 1.

Because of its lack of narrative structure, some of the importance of Part 2 is underplayed. I imagine that, as a member of the September Group, it would be important to Roemer that he has shown how socialism can operate in an exchange economy. His socialism need not rely on altruism or empathy for others in society, rather it relies on a recognition that we are all in the same boat, which leads to solidarity. Indeed, one could see the book as providing a microfoundation for market socialism, with a detailed technical exploration of how that could work. However, it doesn't feel like the book builds up to that crowning achievement.

It may have been unfair of me to focus on the philosophical implications of Roemer's theory, rather than the technical results that are his forte. However, in seemingly writing for an audience of technical economists, Roemer has missed a trick. There is a lot in his work that would be of interest to philosophers and political theorists, if he could explain it in prose. If he ever writes that book, I would love to read it.

REFERENCES

Bacharach, M. 2006. *Beyond individual choice: teams and frames in game theory*. Princeton University Press.

Elster, J. 2017. On seeing and being seen. *Social choice and welfare*, *49*(3-4): 721-734.

Gold, N. 2012. Team reasoning and cooperation. In S. Okasha and K. Binmore (eds). *Evolution and Rationality: Decisions, Cooperation and Strategic Behaviour.* Cambridge: Cambridge University Press

Lewis D. 1969. Convention: a philosophical study. Harvard University Press, Cambridge, MA (Reissued 2002, Blackwell Publishers Ltd, Oxford)

Sugden, R., 1993. Thinking as a team: Towards an explanation of nonselfish behavior. *Social philosophy and policy*, 10(01): 69-89.

BIOGRAPHICAL INFORMATION (maximum 100 words)

Natalie Gold is Senior Research Fellow in the Faculty of Philosophy at Oxford University. Her research area is behavioural decision making, broadly construed; she brings together approaches from philosophy, economics, and psychology. She has published philosophical and empirical research on topics including framing, cooperation and coordination, self-control, trust, and moral judgements and decision-making, including applications to issues in public health and finance.

		Player 2	
		Cooperate	*Defect*
Player 1	*Cooperate*	2, 2	0, 3
	Defect	3, 0	1, 1

(a)
Prisoners dilemma

		Player 2	
		Stag	*Rabbit*
Player 1	*Stag*	3, 3	0, 2
	Rabbit	2, 0	1, 1

(b)
Stag hunt

		Player 2	
		Hawk	*Dove*
Player 1	*Hawk*	0,0	3, 1
	Dove	1, 3	2, 2

(c)
Hawk-dove

Figure 1. Three standard games

		Player 2	
		Dance	*Boxing*
Player 1	*Dance*	3, 2	0, 0
	Boxing	0, 0	2, 3

(a)
Battle of the sexes

		Player 2	
		Dance	*Boxing*
Player 1	*Dance*	3, 2	1, 1
	Boxing	0, 0	2, 3

(b)
Variant battle of the sexes

		Player 2	
		Event I prefer	*Event my partner prefers*
Player 1	*Event I prefer*	1,1	3, 2
	Event my partner prefers	2, 3	0, 0

(c)
Re-described variant battle of the sexes

Figure 2. Three battle of the sexes games