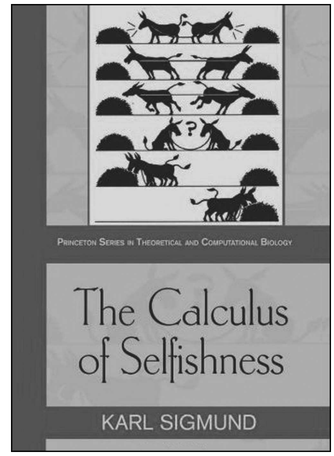
The Calculus of Selfishness

Reviewed by Olle Häggström



The Calculus of Selfishness
Karl Sigmund
Princeton University Press, January 2010
US$37.50, 192 pages
ISBN-13: 978-06911-427-53

Leading game theorist Karl Sigmund calls his latest book *The Calculus of Selfishness*, although arguably *The Calculus of Cooperation* would have been an equally suitable title. The central problem is this: How can cooperation, or even altruism, come about in a population of selfish individuals? We see plenty of cooperation around us, most prominently in human societies but also within other species and even between species. This is a bit of a mystery, because such cooperation appears to be prohibited by Darwinian survival of the fittest, which rewards those individuals who best look after their self-interest. It is they who get to pass their genes (and, presumably, their behavior) on to later generations, as opposed to those who waste time and resources helping others. What is going on here?

A prototype model, mimicking various real-world situations, is the *Prisoner's Dilemma*. Here two individuals, denoted Player I and Player II, simultaneously and independently decide on either of two moves: Cooperate (C) or Defect (D). Each player is rewarded by an amount that depends on both players' moves in such a way that, on the one hand, no matter what the other player does it is better for oneself to play D, whereas on the other hand it is better for both players if both of them play C than if both of them play D.

Olle Häggström is professor of mathematics at Chalmers University of Technology in Gothenburg, Sweden. His email address is `olleh@math.chalmers.se`.

DOI: http://dx.doi.org/10.1090/noti789.

A special case is the *Donation Game*, which is used by Sigmund as an example throughout much of the book. Here each player can choose to donate \$5 in order for the other player to receive \$15 (C) or to refrain from donating (D). Obviously it is better for both players if they both donate, leaving each with a net benefit of \$10, than if they both refrain, in which case they get nothing. On the other hand, each player has the incentive to save \$5 by refraining, no matter what the other player does. Hence, it seems that two self-interested players are doomed to both play D, thus missing out on the \$10 benefit that each of them might otherwise have acquired.

There are various ways to try to explain how the players might nevertheless come to play C. One approach, sometimes favored in evolutionary biology, is kin selection. Helping a sibling can be in my interest—or rather in the interest of my genes—because she shares 50% of them. From this perspective, the \$5 donation is a bargain, because the \$15 that my sibling receives is worth \$7.50 to me.

Another approach, receiving more attention in Sigmund's book, is repeated games. If you and I are set to play the Prisoner's Dilemma many times, then it might start to look like a good idea for me to play C early on, with the idea of building up a relation of trust in which you are more inclined to play C in later rounds than you otherwise might have been (and vice versa). The situation quickly becomes incomparably more complex than in the single-round game, due partly to a combinatorial explosion in the number of possible strategies.

Robert Axelrod's classic 1984 book *The Evolution of Cooperation*, which did much to popularize the subject and to stimulate further research, reports on a fascinating experiment. He invited colleagues and others to submit computer programs to play the iterated Prisoner's Dilemma with each other. After a first round-robin tournament, he published an analysis of the results and opened invitations for another one.[1] A wide variety of strategies were submitted, but remarkably enough both tournaments were won by the same very simple strategy, called *tit for tat*, which plays C in the first round and from then on simply copies what the opponent just did. Tit for tat is by no means a universal winner independently of the environment of co-competitors—for instance, in his analysis of the first tournament Axelrod gave an example of a variation of tit for tat that would have won (ahead of tit for tat and all others) if only it had been submitted. Still, tit for tat performs well under sufficiently wide conditions to merit the large amounts of attention it has received in the game theory literature, including Sigmund's book.

All this is an example of game theory, which can be described as decision theory in the presence of other agents whose decisions affect how successful your own decisions are (and vice versa). This is interesting both as a mathematical topic in its own right and for modeling in biology and in economics. One of several aspects that contribute to game theory being a fascinating subject is the multiplicity of methodologies involved. In current research we find rigorous mathematical analysis, we find extensive computer simulations, and we find experiments aimed at uncovering how real human beings act in idealized game-theoretic situations. Sigmund concentrates on mathematical analysis, but there is of course interesting interplay with the other approaches, which he does not ignore.

Of particular interest is *evolutionary game theory*, in which one imagines a large population of agents playing with each other and in which the frequency of a given strategy in the population changes depending on how successful it is. With a symmetric game with n possible strategies, let A be the *payoff matrix* of dimension $n \times n$, where A_{ij} denotes Player I's payoff if he plays strategy i and Player II plays strategy j. Furthermore, let $x_i(t)$ denote the frequency at time t of strategy i in the population, and $\mathbf{x}(t) = (x_1(t), \ldots, x_n(t))$. The so-called *replicator equation* prescribes that the rate of relative change of $x_i(t)$ equals the success of a player with strategy i in the population minus the average success in the population:

$$x_i'(t) = x_i(t)[(A\mathbf{x}(t)_i) - \mathbf{x}(t) \cdot A\mathbf{x}(t)]$$

for $i = 1, \ldots, n$. In biological applications, we may think of this as describing the change of gene frequencies in the population resulting from Darwinian selection, whereas in economics it may be more natural to think of it as resulting from agents imitating the behavior of other, more successful, agents.

The replicator equation leads to fascinating and not-so-easy-to-guess dynamics even in seemingly simple situations. An important notion here is that of *rest points*, that is, population compositions such that $x_i'(t) = 0$ for $i = 1, \ldots, n$. A central question is whether they are stable under perturbations. Rest points are closely related (but not quite equivalent) to so-called Nash equilibria. The bulk of Sigmund's book consists of analyses of the replicator dynamics in various situations.

One example is the iterated Prisoner's Dilemma, which, under reasonable conditions, can be represented as a single game and plugged into the replicator equation. There are infinitely many possible strategies for this game, and we need to focus on a finite collection of them. Already with three simple strategies—tit for tat (TFT), always cooperate (AC), and always defect (AD)—the dynamics become fairly intricate. If only TFT and AC are present in the population, everyone will cooperate forever, and the two kinds of agents will look the same to the outside observer. Introducing a noise term in the dynamics, allowing TFT and AC to mutate into each other, causes the frequencies of the two strategies to diffuse back and forth. What if we also introduce another mutation term, allowing occasional attempts by AD to invade the population? If, at the time of an invasion attempt, the population is dominated by AC, then AD will be able to exploit AC and quickly take over the entire population. On the other hand, if it is dominated by TFT, then AD will fail and be eliminated, at the same time as the proportion of TFT will increase at the cost of AC. This has the interesting consequence that, starting from a population dominated by TFT plus a smaller proportion of AC, AD will have a very hard time invading if the mutation rate in its favor is too high. This is so because, every time it attempts to invade, the proportion of TFT will go up, making the next attempt even less likely to succeed. Only if invasion attempts come more rarely, so that the TFT-versus-AC proportion has time to diffuse towards a significantly greater proportion of AC between invasion attempts, does AD have a chance to succeed in taking over the population reasonably quickly.

There are many natural ways to vary this situation. Introducing an error term in the players'

[1] *In this second tournament, he also adjusted for a flaw in the first one, namely that the number of iterations of the games was fixed beforehand at 200, opening up the nightmare of backward induction: Clever contestants realize that they have no reason to play anything other than D in the last round. It follows that they have no such reason in round 199, either, and so on, seemingly leading rational players to defect from round 1. In the second tournament the number of rounds between each pair of contestants was announced to be random—more specifically, geometrically distributed with a given expectation.*

choices (so that they sometimes make a different move than intended) can alter the dynamics drastically and tends to disfavor tit for tat. Another option, particularly popular in recent years, is the study of *indirect reciprocity* (as opposed to the direct reciprocity of tit for tat in the iterated Prisoner's Dilemma), in which two agents meet only once but can nevertheless adjust their moves depending on what has taken place before. This can be surprisingly effective, provided the agents apply *vicarious reciprocity* (Player I cooperates with Player II depending on how Player II has previously behaved towards third parties) rather than the psychologically tempting *misguided reciprocity* (Player I cooperates with Player II depending on how Player I himself has previously been treated by third parties).

Yet another variation treated by Sigmund is a multiplayer generalization of the Prisoner's Dilemma, sometimes known as the *Tragedy of the Commons*. The commons is a piece of grassland owned collectively by a group of local farmers. Everyone is free to use it for their sheep, but the tragedy consists in the fact that, if the total level of exploitation of the commons gets too large, it will collapse. The importance of understanding a game like the Tragedy of the Commons is its structural similarity with many important problems—environmental and others—in society; finding ways to stimulate agents to cooperate in Tragedy of the Commons-like situations is an enormously important task for political science and related subjects, and one may hope that game theory is able to contribute. A major example is the emission of greenhouse gases, where I find it disheartening to see that the biggest polluter[2] seems to be dead set on continuing to defect.

As a prerequisite for the book, a year of undergraduate mathematics should be more than sufficient. More important, however, is that the reader is willing to focus seriously on the material and to invest time and energy. Sigmund's enthusiasm for the topic shines through very clearly, and I enjoyed reading this up-to-date introduction to a very lively and exciting research area. Nevertheless, I would hesitate to use the book for a course. The reason for this is Sigmund's choice, despite the mathematical rigor he employs, to settle for a narrative structure entirely devoid of the definition-lemma-theorem-proof-exercise layout that, after all, has proved over the years to be an efficient means by which to structure and communicate mathematical ideas.

[2] *I am referring to the United States. Although it is true that, in absolute terms, China has recently surpassed the United States as the biggest emitter of greenhouse gases, it is nevertheless the case that the United States is still way ahead of China in terms of per capita emissions, as well as in terms of cumulative (historical) emissions.*