



© 1997–2004, Millennium Mathematics Project, University of Cambridge.

Permission is granted to print and copy this page on paper for non-commercial use. For other uses, including electronic redistribution, please contact us.

Mar 2002

Regulars



Mathematical mysteries: Survival of the nicest?

by Helen Joyce



Survival of the nicest?

One of the most puzzling aspects of human behaviour is cooperation, in situations where backstabbing and selfishness would seem to be more rewarding. From the point of view of evolutionary theory, the very existence of altruism and cooperation appear mysterious. The mechanics of evolution seem to imply that rugged competition is the order of the day; that, given an opportunity to benefit by cheating someone, or by defaulting on a deal, we will inevitably do so. Surely to do otherwise would mean relegation to the sidelines of the evolutionary game – and in that game, demotion means extinction.

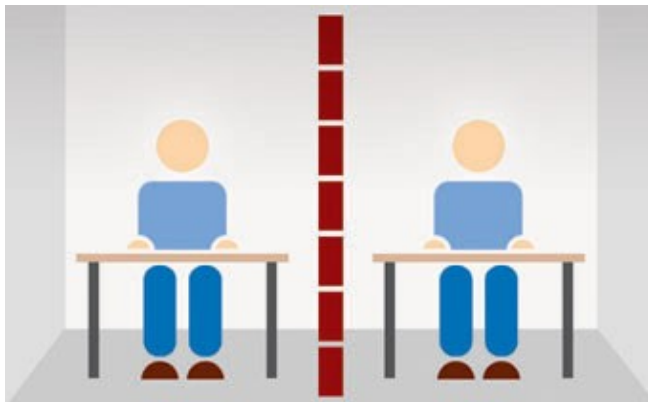
In fact, as even the most cynical observer must admit, cooperation is rife in human society. Even if you sometimes despair of human nature, you must admit that the "dog-eat-dog" scenario conjured up by the phrase "survival of the fittest" doesn't bear much resemblance to life as we know it. So it must be that, from a purely selfinterested point of view, cooperation can actually be good for us.



Solitary confinement

To highlight this puzzle, consider the Prisoner's Dilemma, described in detail in [Adam Smith and the invisible hand](#) in Issue 14 of *Plus*. To summarise, this very famous paradox in Game Theory describes two people suspected of being accomplices in a crime. They are held prisoner in separate, non-communicating cells. The police visit each prisoner, and tell both that if neither confesses, each will be sentenced to two years in jail. However, if only one prisoner confesses, implicating the other, the one who confesses will get off scot-free as a reward, and the other, who didn't confess, will receive a punitive sentence of seven years. If each confesses and implicates the other, both will be sentenced to three years.

What should a prisoner do in this situation? Suppose the other prisoner doesn't confess. Then the best course of action is to confess, and go free. Even if the other prisoner does confess, it will be better to have done likewise – at least the sentence will be lower. Both prisoners will reason thus, so both will confess and end up serving sentences of five years – even though, if both had remained silent, both would have served sentences of only three years.



Second-guessing

If you think this dilemma is very far from your everyday life – after all, you are lawabiding and will never be thrown in jail! – think again. Every time you make a bargain, you are potentially facing the prisoner's dilemma. What is to stop you – or, more to the point, the person you are making the bargain with – from defaulting? Surely both of you will be tempted by the prospect of getting something for nothing, and afraid that if you are honest the other person won't be, and you'll get landed with the so-called "sucker's payoff" – getting nothing for something? It's all very well to say that "honesty is the best policy" but surely this is a

Mathematical mysteries: Survival of the nicest?

luxury that only the civilised and comparatively rich can afford?

Well, the good news is that we are not dependent on the benevolence of others, as the prisoner's dilemma would seem to suggest. In fact, cooperation can spontaneously break out even among fundamentally selfish agents – provided you assume that people meet each other more than once, and can remember what the other person did last time they tried to strike a bargain.

To explore this sort of situation, political scientist Robert Axelrod invented the game of "Iterated Prisoner's Dilemma" – Prisoner's Dilemma played repeatedly against the same opponent – and set up a tournament, inviting academics from all over the world to devise strategies. First Axelrod compared various strategies by pairing them and seeing who won; then he held a meta-tournament, in which there were many agents, each with its own strategy which it was allowed to modify in response to what was going on around it, for example if it saw that other agents had more successful strategies.

Over the long term, Axelrod discovered that selfish strategies tended to do very badly, as did foolishly generous strategies. Defecting encouraged others to defect; not punishing others for defecting only encouraged them to do so again. One of the most successful and stable (in other words, successful against many different strategies and in many different environments) was "Tit for Tat". This strategy involves cooperating the first time you meet another agent, and after that always repeating your opponent's last move. So if your opponent defaults on one turn, you punish them by defaulting on the next; if they cooperate on one turn, you reward them by cooperating on the next.

A slightly better strategy – because it avoids the possibility of getting trapped into long cycles of retaliation – is "Tit for Tat with forgiveness". This is Tit for Tat with a small randomised possibility of forgiving a defaulter by cooperating anyway. Forgiveness is particularly helpful if you introduce the possibility of misinformation into the game – that is, if moves are sometimes randomly misreported.

The submitted strategies varied in many ways – initial hostility, tendency to forgive or retaliate, complexity, how much past behaviour they took into account, and so on. No one strategy will always be best because how a strategy does depends on who the other players are – if you're playing against mugs, you may as well be a freeloader, and if you're playing against sharks, you may as well get your retaliation in first! And research into human behaviour is ongoing, with biologists, economists and mathematicians studying phenomena such as spiteful behaviour, altruism, and kin selection (generosity between close relatives, which is evolutionarily useful since their genes are similar). But Tit for Tat did well or best in Axelrod's tournament against very many different opponents – showing how cooperation could evolve using only the selfish mechanisms of natural selection.

About the author

[Helen Joyce](#) is editor of *Plus*.



Plus is part of the family of activities in the Millennium Mathematics Project, which also includes the [NRICH](#) and [MOTIVATE](#) sites.