Evolution of Cooperation Through Similarity Recognition

Project funded by the Nuffield Foundation, January–December 2009

Andrew Colman  Briony Pulford  Lindsay Browning
Investigator  Investigator  Researcher

Background

Cooperation is usually formalized in terms of the Prisoner’s Dilemma game, shown below, and the problem is how to explain C-choosing in this Prisoner’s Dilemma game:

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>3, 3</td>
<td>0, 5</td>
</tr>
<tr>
<td>D</td>
<td>5, 0</td>
<td>1, 1</td>
</tr>
</tbody>
</table>

Player I chooses between rows C (cooperate) and row D (defect), Player II chooses between columns C and D, and the pair of numbers in each cell are the payoffs to Players I and II respectively in the corresponding outcome. Each player does better by choosing D than C irrespective of what the co-player chooses; the D strategy is the unconditionally best or dominant strategy for both players. The puzzle arises from the fact that if both players choose D, as they are bound to if they are rational, then each is worse off than if both choose C, and many people therefore have a
strong intuition that cooperation makes sense. Furthermore, whenever the behaviour of real decision makers is studied in experimental games, rampant cooperation is observed (Colman, 1995, 2003; Sally, 1995).

Many two-person interactions involving cooperation and competition, trust and suspicion, threats, promises, and commitments are Prisoner’s Dilemmas or multiplayer versions of it. A familiar example is overfishing cod in the North Sea. It is in the self-interest of each fishing company to catch as many cod as possible, whether or not others are restraining their catches, but if all behave in this way, then the cod are fished to extinction, and each fishing company is worse off than if they had all cooperated by exercising restraint. Cooperation is frequently observed in everyday life, but it is difficult to maintain when individuals are tempted to defect: herring were fished to near-extinction round Britain in similar circumstances in the early 1970s.

It is notoriously difficult to justify C-choosing in the Prisoner’s Dilemma game. Binmore (1994, chap. 3) reviewed many naive attempts. However, if two players are close relatives, then both maximize the number of their own genes that they transmit to the next generation by cooperating, and this may explain the evolution of alarm calls and similar cooperative and altruistic behaviour through kin selection or inclusive fitness (Hamilton, 1964). Among unrelated players, if a game is repeated an indefinite number of times, then reciprocal altruism (Trivers, 1971) – I’ll scratch your back if you’ll scratch mine – may evolve as a sensible strategy promoting cooperation. In non-repeated interactions, indirect reciprocity (Alexander, 1987) can explain the evolution of cooperation, provided that the interactions are observed by others with whom future interactions might occur. What is difficult to explain is cooperation between unrelated individuals in unrepeated interactions lacking opportunities for reputation management, although it is often observed in everyday life.

An interesting and persuasive mirror strategy explanation, based on similarity recognition, was put forward by the operational researcher John Howard (1988), and independently by the philosopher Peter Danielson (1992). Both proved that players who are completely rational have a reason to cooperate if they recognise that their co-players are identical to themselves. Both formalized this in terms of game-playing automata that can compare their programs and recognize identical programs. Howard implemented the argument in Basic, and Danielson implemented it in Prolog. The underlying idea can be traced back to Gauthier (1986, and earlier articles). Informally, the argument is that each player, knowing that the co-player is identical, can reason validly that any strategy chosen will also be chosen by the co-player, because the co-player is literally identical and in the same situation and must therefore make the same choice, and it is therefore rational to choose C in Figure 1 (and earn 3) rather than D (and earn 1). Furthermore, a population of such players would be evolutionarily stable against invasion by purely selfish D-choosers.

This is a valid proof, as its computational implementation proves, but its relevance to cooperation in human and animal populations is limited by the requirement that players have to be identical (and able to recognize their similarity). Although the mirror strategy argument applies only to identical clones, strictly speaking, it is worth noting that female Hymenoptera (bees, ants, wasps), which very unusually share 75% (rather than 50%) of their genes in common, are exceptionally cooperative and
self-sacrificing, although neither Howard (1988) nor Danielson (1992) mentioned this. We plan to investigate this phenomenon further, using a novel methodology. The proposed research focuses on a radical approach to an old problem.

References


<back to Andrew Colman's home page>