The Moral Herd: Groups and the Evolution of Altruism and Cooperation

Garcia Gallego, J.A.

2011

document version
Publisher's PDF, also known as Version of record

citation for published version (APA)
Garcia Gallego, J. A. (2011). *The Moral Herd: Groups and the Evolution of Altruism and Cooperation*. TI/Thela Thesis.

Economic science assumed for long that we, humans, are selfish and do not care about others. Recently we have started to find out that the real face of human helping behaviour is not that dark, but also not that simple. Our species is generally *born and bred* to be helpful, as attested by our very apparent capacity for collective action. But we are also prepared to be selfish and even harm others in specific contexts. In other words, our helping behaviour is complex. In this thesis I examine the origins of humans' taste for helping behaviour by looking precisely at those complexities in four different essays.

In the first essay (chapter 2) we study cooperation in large groups of unrelated individuals. We introduce a model that allows for the co-evolution of two related traits. One trait determines individuals preferences for size of the group in which interaction takes place; the other trait determines how cooperative individuals are. Groups are disbanded and reformed every generation so that interactions take place between unrelated individuals.

Using a two-trait optimization approach, we derive analytical solutions for the levels of cooperation and group sizes that are expected to arise from the joint evolution of these two traits. Simulations show that these predictions generally hold when including stochastic effects and heterogeneity. The formation of groups and the tendency to cooperate are shown to interact in ways that highlight the importance of looking at the two processes jointly. The dynamics bring the population to a point at which the marginal costs of the traits are zero, that is, agents reap no benefits by marginally raising their grouping tendencies or cooperation levels.

Equilibria in the model reflect cases where cooperation can be interpreted as the provision of a public good. The fitness values are the payoffs that accrue to individuals in the same group for investing in a common project. In groups of two or more individuals it is impossible to exclude those contributing less than the average from the benefits of the common project. Since positive contributions arise in different equilibria with large groups, we are able to explain the provision of public goods without invoking relatedness.

In chapter 3 we set out to explain why cooperation often discriminates against out-groups. We study the evolution of discriminatory strategies in a prisoner's dilemma game considering the interaction between four strategies; namely egoists, altruists, parochialists and traitors. Intergroup interactions harm the evolution of cooperation. Since early humans may have had few or many intergroup contacts, the ability to discriminate individuals from other groups is a possible mechanism that would allow co-operators to sustain helping inside their groups.

We develop a computational model with selection occurring at the levels of groups and individuals. Three versions allow us to test different hypothesis on how parochial strategies originated: namely, in which selection is exclusively driven by assortment, exclusively driven by group conflict, and driven by both.

Both group conflict and assortment select for parochial behaviour. Selection for such behaviour is strongest when the two selection mechanisms are both present, migration is low, and chance has a small influence on winning a conflict. In all of the studied settings helping behaviour (inside the group) always comes with hostility to- wards outsiders. This would seem to suggest that the advantages of helping often come with the disadvantages of discrimination.

In chapter 4 we study the evolution of strategies in repeated games. Repetition has often been considered as crucial for human cooperation, but is usually studied using restricted strategy sets. We focus on a simple setting that captures the essential features of repetition without restricting the strategy space. The case study addresses the evolution of reciprocal cooperation in a repeated prisoner's dilemma. We use the standard tool-set from evolutionary game theory as well as computer simulations to obtain results.

The results indicate that instability is intrinsic to the repeated prisoner's dilemma in the presence of a rich strategy space. Every equilibrium can be upset, either by a mutant or by a succession of mutants. Under very reasonable mutation schemes these stepping stone paths in and out of equilibrium not only exist, but evolution also actually finds them. Whether we can expect cooperation to increase or decrease depends on how often indirect paths in and out of cooperation are induced by mutations. In other words, mutations should not be ignored and their specification matters.

In chapter 5 we study the interaction of repetition and population structure. These two mechanisms are prevalent when discussing human cooperation. Economists often emphasize the role of repetition, while biologists seem to stress more the role of population structure. To our knowledge, the two mechanisms have never been studied together.

We develop a model with two main parameters: the level of assortment and the continuation probability in a repeated game prisoner's dilemma. The former controls population structure, while the latter reflects the role of repetition. By analysing pair-wise stability in a restricted strategy set we are able to classify the parameter space of the model into four regions. Each corresponds to a different predicted level of cooperation. Humans arguably belong to a region in the parameter space where the continuation probability of the repeated game is large, and assortment is small but positive. This region contains cooperative equilibria only, but different equilibria here have different levels of cooperation.

The combination of small but positive levels of assortment and large continuation probabilities seems to go very well with the nature of our social preferences. Reciprocity is high when the continuation probability is high, which matches the human taste for reciprocity. Assortment is able to explain cooperation in one-shot interactions. In the presence of repetition, a marginal change in assortment goes a long way in promoting cooperative behaviour.

A common ground for all the models proposed in this thesis is that strategies are not too simple. In chapter 2 individuals choose a level of cooperation as well as a grouping tendency that will determine the size of their groups. In chapter 3 strategies are flexible in the sense that helping can vary depending on who individuals are inter- acting with. Chapters 4 and 5 consider a rich set of strategies that takes into account the behaviour of the opponent when the game is repeated. All these specific features (variation in cooperative behaviour according to group size, discrimination and reciprocity) reveal themselves in experimental settings. A satisfactory theory of human cooperation should therefore explain how all of these, and other experimentally relevant features originated. This suggests that it is likely that one single approach will fail to deliver the whole picture for the human cooperation puzzle. It would therefore be good to devote more attention to explaining how existing models relate to each other.