On the origins of human sociality

Akdeniz, A.

Publication date
2023

Chapter 1

Introduction

Human nature contains both good and evil. We observe displays of extreme prosociality and substantial cruelty in human societies around the world. To answer the question of how these two extremes can co-exist, and why we observe any prosociality at all, given its costs, I study the origins of human sociality. If we understand where our social preferences and behaviours come from, we can then understand why people do or do not care about one another, and why there exist high levels of heterogeneity in sociality across societies, across individuals within a society, and across time and space within an individual.

My dissertation, therefore, centres around exploring the origins of and the variation in human sociality. In my projects, I study how humans evolved to be ultra-social while the extent of sociality in other animals, and especially in non-human primates, remained relatively restricted. In doing so, I focus on the mechanisms that can explain the co-existence of good and evil in us, and look for unique elements in human social interactions, or a combination thereof, to pinpoint the cause of divergence in human sociality.

Chapter 2 examines a critical assumption from the group selection literature. Group selection models combine selection pressure at the individual level with selection pressure at the group level. Cooperation can be costly for individuals, but beneficial for the group, and therefore, if individuals are sufficiently much assorted, and cooperators find themselves in groups with disproportionately many other cooperators, cooperation can evolve. The existing literature on group selection generally assumes that competition between groups takes place in a well-mixed population of groups, where any group competes with any other group equally intensely. Competition between groups however might very well occur locally; groups may compete more intensely with nearby than with far-away groups. We show that if competition between groups is indeed local, then the evolution of cooperation

can be hindered significantly by the fact that groups with many cooperators will mostly compete against neighbouring groups that are also highly cooperative, and therefore harder to outcompete. The existing empirical method for determining how conducive a group structured population is to the evolution of cooperation also implicitly assumes global between-group competition, and therefore gives (possibly very) biased estimates.

Chapter 3 experimentally tests the role of partner choice in the evolution of honesty. There are many situations in which the ability to lie can help one get ahead. Most people however hesitate to lie and regularly prefer to tell the truth, even if lying could work to their advantage. Why honesty exists therefore is an open question. We explore the possibility that it has evolved because of partner choice. In a lab experiment, we find that in a situation in which their partner will know more than they do, subjects do prefer to be matched with honest partners. We also find that they are right in doing so because honest partners will behave consistently, and on average more, prosocially. We do not find support for the possibility that honest partners behave more prosocially in order to avoid the choice between lying and revealing having been selfish. Instead, our findings are consistent with an explanation in which honest people behave more prosocially because they also have a harder time justifying selfish behaviour to themselves.

Chapter 4 presents an extensive literature review, where we compare the explanations for human cooperation from the theoretical literature with the empirical observations from the experimental literature, and argue that there is a mismatch between the two. A considerable share of the literature on the evolution of human cooperation considers the question of why we have not evolved to play the Nash equilibrium in prisoners' dilemmas or public goods games. In order to understand human morality and pro-social behaviour, we suggest it would actually be more informative to investigate why we have not evolved to play the subgame perfect Nash equilibrium in sequential games, such as the ultimatum game and the trust game. The "rationally irrational" behaviour that can evolve in such games gives a much better match with actual human behaviour, including elements of morality such as honesty, responsibility, and sincerity, as well as the more hostile aspects of human nature, such as anger and vengefulness. The mechanism at work here is commitment, which does not need population structure, nor does it need interactions to be repeated. We argue that this shift in focus can not only help explain why humans have evolved to know wrong from right but also why other animals, with similar population structures and similar rates of repetition, have not evolved similar moral sentiments. The suggestion that the evolutionary function of morality is to help us commit to otherwise irrational behaviour stems from the work of Robert Frank (1987; 1988), which has played

a surprisingly modest role in the scientific debate to date.

Chapter 5 examines whether deviations from selfishness in the ultimatum game can be explained by noise or mistakes in individuals' choices. In this chapter, we review, upgrade, and synthesize existing models from evolutionary game theory on the ultimatum game, and we compare their predictions with the existing experimental evidence. We find that the results in Gale et al. (1995) and Rand et al. (2013) are primarily driven by bias in the mutations. We make versions with local instead of global mutations for both. This minimizes the bias and changes the results. We also consider Quantal Response Equilibria in combination with the assumption that individuals are selfish after all. The Quantal Response Equilibrium is the noisy twin of the Nash equilibrium, and looking at this combination we explore an alternative explanation for what we observe in the lab, namely noise instead of deviations from selfishness. Finally, we provide a refurbished version of the model of commitment in Nowak et al. (2000). The de-biased version of the model in Rand et al. (2013) becomes a special case of this more general model (with the possibility for commitment muted). We find that the experimental evidence does not align with the models in Gale et al. (1995), Rand et al. (2013), or our de-biased versions of them, and that it also rejects the combination of selfishness and the Quantal Response Equilibrium. All of these models predict that the distribution of minimal acceptable offers should start with high frequencies at 0, end with low frequencies at 1, and have decreasing frequencies in between, which is not what is found in lab experiments. Instead, the experimental evidence is in line with a commitment-based explanation, where the ability to commit to rejecting unfair offers, while being *ex-post* suboptimal, can be *ex-ante* beneficial.