UiT
THE ARCTIC
UNIVERSITY
OF NORWAY

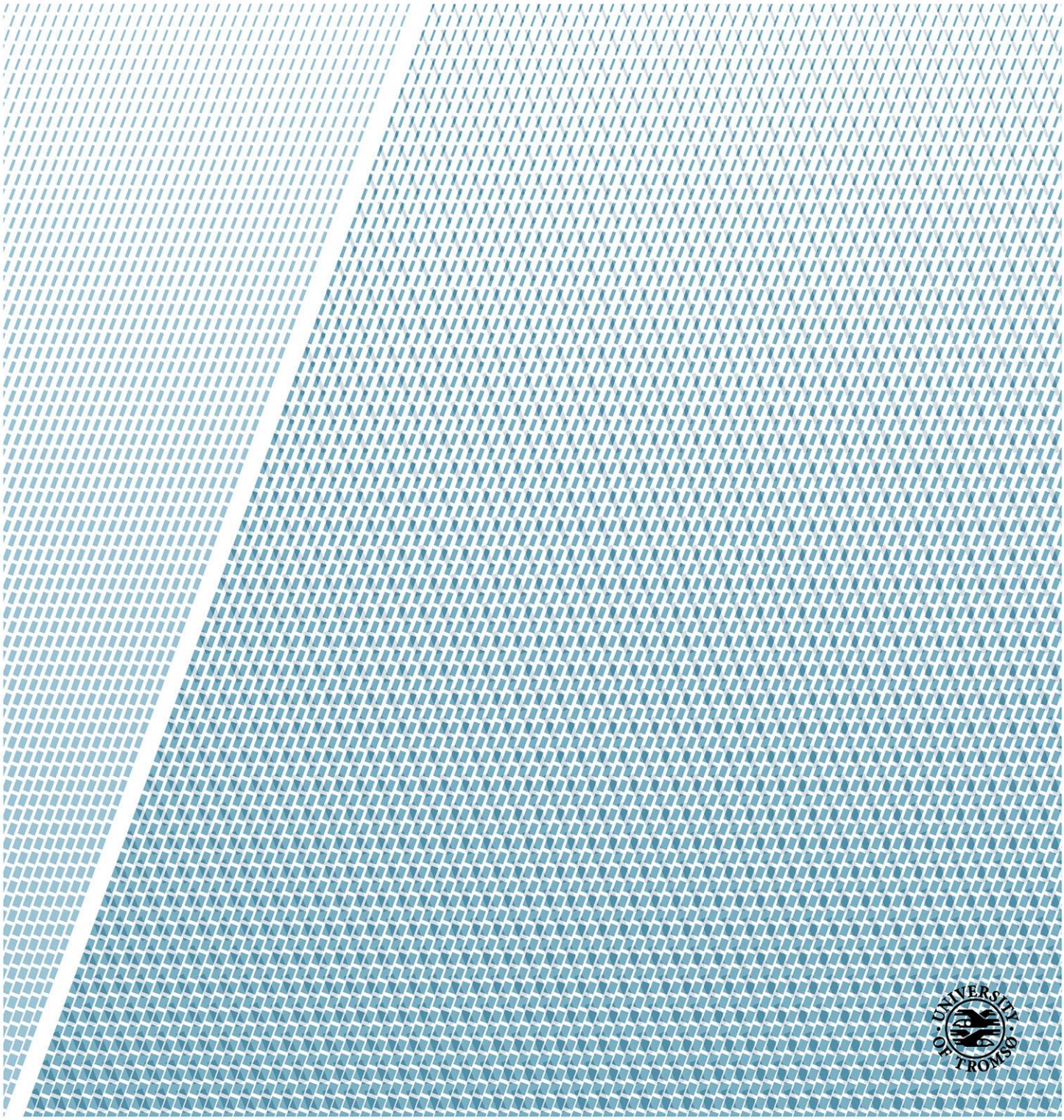
Department of Artic and Marine Biology

Human teenagers defect more often when told to be pure altruistic

—

Ola Sandsdalen

Master's Thesis in Biology and Education BIO-3906 - May 2016



Abstract

Humans (*Homo sapiens*) are known to cooperate with non-kin and even strangers, yet there is no absolute explanation to human cooperation nor to non-kin altruism. Specific human traits such as memory and reputation fuels systems of reciprocity, policing, reward and punishment which all contribute to cooperation in social interactions. How do people adopt their moral heuristics to knowledge of the most common beneficial way to act in confrontations with others? In this experiment I test this by anchoring teenagers to different strategies, either pure altruistic or altruistic punishing, before playing iterate prisoner's dilemma. In contrast to findings in similar experiments, no differences in player's total cooperative plays, score nor mutual cooperative plays was found in this study. Yet, when anchored with pure altruistic strategies participants adopted nastier strategies than participants anchored with strategies for altruistic punishment. That is, I found no difference in their forgiveness, but free riding tendencies were more common in the groups anchored for pure altruism. In sum, human teenagers do not seem to get nicer by being told that it's in their best interest to be nice - rather the opposite.

1.Introduction

Altruistic behaviour can be defined as beneficial behavior towards another not closely related individual, while simultaneously being detrimental to the individual performing the behavior. Benefit and detriment is here defined in terms of contribution to inclusive fitness (Trivers, 1971). Costly and fitness reducing acts which benefits other individuals are to a great extent restricted to kin groups, but in several species individuals show altruistic tendencies towards unrelated individuals (Fehr and Fischbacher, 2005). Yet, there is no absolute explanation for this behavior and some solutions have been proposed; reciprocity (Trivers, 1971; Axelrod and Hamilton, 1981), policing (Ratnieks, 1988; Frank, 1995) and systems of reward or punishment (Rapoport and Guyer, 1966; Oliver, 1980; Fehr, 2002).

When repeated interactions are possible, the rewards for cooperation can be larger than the short-term pay off of a single defection. A classic example of this is found in vampire bat (*Desmodus rotundus*) colonies, where individuals are observed to share blood-meals with unrelated neighbours (Wilkinson, 1984). For the recipient, the meal could potentially save its life, while the donor, after a successful hunt for a large and resource rich blood meal, has little cost of giving away a small quantity of its catch. Given a stable neighbourhood and repeated interactions, reciprocal altruistic interactions like these may be stable in long term interactions (Trivers, 1971). Yet, the

disadvantage for reciprocal altruists is their vulnerability towards individuals who are happy to accept help, but also eager to "forget" the payback. These non-reciprocal individuals (or free riders) reduce the fitness of altruists and they may therefore make reciprocity less likely to evolve (Axelrod and Hamilton, 1981; Alcock, 2009).

In humans, genetically unrelated individuals cooperate within large groups and even with total strangers from other cultures, with everything from warfare to food-sharing (Nowak and Sigmund, 2005; Boyd and Richerson, 2005). Building a reputation, which is easily shared through language, is one of the key aspects that separate humans from other animals, and the ability to recognize and remember cheaters is a trait inherent to humans (Dunbar, 1993; Milinski et al., 2002). It has been suggested that humans are very attentive to possibilities of individual reputation building, and when given the opportunity to gain a reputation for generosity, helping rates increase strongly (Fehr and Fischbacher, 2005). Additionally, in decisions of whether to engage in reciprocity towards other individuals or not, positive or negative reputation has a strong influence. On the proximate level, certain brain signals are shown to promote cooperation towards partners with positive reputation, and these signals are absent towards partners with negative reputation, leading to more defection towards the latter (Phan et al., 2010). Therefore, the probability of repeated interactions are important, as anticipation of re encountering a present partner increases the chance of present cooperation (Gachter and Falk, 2002; Van Lange et al., 2011).

Punishment, on the other hand, provides a secondary tool to deprive free riders, if those who exploit others cooperative actions are punished frequently enough (Heinrich and Boyd, 2001). Yet, there is often no direct fitness gain for the punisher. Thus, from the view of the entire group, punishing (i.e., altruistic punishment) of free riders is beneficial, but the incentive to punish from the individuals point of view is not (Fehr and Gachter, 2002). Punishment as a costly signal of trustworthiness could however benefit individuals in reciprocal groups, as humans are shown to rather reciprocate with individuals that have a reputation of altruistic punishment, than with individuals who don't punish (Kiyonari and Barclay, 2008; Jordan et al., 2016). Humans are also able to use logic thinking to foresee outcomes of future interactions, so that choices made today can determine later choices of the players (Axelrod, 1984). That is, subjects often cooperate because they foresee the rewards for cooperation and the punishments for non-cooperation, and therefore adopt a longer-term perspective for the present situation (Axelrod, 1984; Van Lange et al., 2011).

Models of game theory has been popular tools for modelling individual interactions in

species from bacteria to humans. In order to understand reciprocal altruism in humans, a range of games have widely been used, including both two-player and multi player games with iterate and finite durations using different treatments (one-shot and repeated), and the published papers are in the thousands (Maynard Smith, 1982; Dugatkin and Reeve, 1998). Although papers dealing with two-player interactions have been criticized for being unrealistic, as interactions in real life always consist of multiple interaction, they can still be considered a special case; a small part of a larger multiverse of interactions (Gokhale and Troulsen, 2010).

Prisoner's dilemma is a two player binary choice matrix, non-zero sum game that has been widely used as a model in biological- and social-sciences (Rapoport and Cammah, 1965; Maynard Smith, 1982; Axelrod, 1984). The game gives two choices of action; cooperate or defect. For any round of the game, defection yields a higher pay off than cooperation independent of the choice of opponent. The dilemma is that since defection has the highest potential benefit and cooperation the highest potential risk, the equilibrium of the Prisoner's Dilemma is mutual defection. This equilibrium is deficient because the best outcome for both players is mutual cooperation (Rapoport and Cammah, 1965; Maynard Smith, 1982; Axelrod, 1984). Studies of iterate games of prisoner's dilemma often emphasize participant's strategies (or decision rules), where strategies are specifications of what to do in any situation that may arise. How subjects design their strategies depend on several factors, from probabilities, past plays and previous patterns (Axelrod, 1984). Through extensive computer simulations, cooperative and forgiving strategies have been shown to outperform strategies relaying on defections (Axelrod, 1984; Nowak and Sigmund, 1992).

How human behaviour in iterated rounds of prisoner's dilemma is influenced by individual players immediate prior stimulus (ie., anchoring) was first studied in the late 1950's. Deutch (1958) induced players to feel cooperative or individualistic prior to playing one-shot and iterate games of prisoner's dilemma, in both treatments players induced to feel cooperative, indeed cooperated with their partner to a greater extent than those who received the individualistic treatment. It is also documented that subtler contextual manipulations than those of Deutch (1958) can promote strategic changes in players, promoting either enhanced cooperative or defective plays both in one-shot and iterate games (Ellingsen et al., 2012). Anchoring (i.e., priming or framing) effects in general has been studied mostly through numerical stimulus before solving informative or uninformative tasks (Tverksy and Kahneman, 1974). The term is described in various ways, but it is mainly known as the effect of a specific stimulus on the psychophysical judgment of another stimulus. A known metaphor for anchoring is a boat using its anchor to keep it stable at rough sea,

in the same way we can use mental anchors as a holding point in the sea of thought. An implemented idea or thought can help judgment in new tasks, for example, anchoring of a stimulus like a value, will affect the judgment of deciding the numerical value of a task (Chapman and Johnson, 1994). All anchoring procedures involves presentation of an anchor and it is shown that anchors do have an effect on judgment and reasoning, either they are informative or uninformative (Tversky and Kahneman, 1974; Green et al., 1996).

In this study informative and instructive anchors are presented, and it is expected that participants would incorporate these anchors in their strategies when playing prisoner's dilemma (Chapman and Johnson, 1994). Therefore, I examined whether instructing individuals to either be altruistic punisher's or pure altruists had effects on individual behaviour playing prisoner's dilemma. That is, how do individuals incorporate the external stimulus in their strategic choices playing the game? Given anchoring effects, it is expected that subjects would use anchors and incorporate them into their strategy.

2.Materials and methods

2.1 Ethics statement

All participants in the experiments reported, signed an informed consent to participate. Additionally, their anonymity was always preserved (in agreement with the Norwegian Law for Personal Data Protection) by assigning them as player 1, player 2 and judge within random groups. No association was ever made between their real names and their results.

2.2 Methods

The experiment was conducted in 6 high school classes of 24 - 28 students each (N = 94). Subjects where either freshmen or second-year's at high school with a mean age of 16.5 ± 0.5 years and with a slight female biased sex ratio (56 %). For all groups, practical implementation of plays where held in their classroom with both their science teacher and the researcher present. Every subject where given a formal introduction to the rules of iterated prisoner's dilemma and introduced to the pay out matrix (appendix 1). The subjects where then divided into groups of three, with one judge and two players according to the seating diagram (appendix 2). Within each class, subjects could choose their own groups to their liking, resulting in groups consisting almost entirely of well

known acquaintances (pers. obs.). Each player where given two play cards designated “cooperate” and “defect” to keep stored on their thighs. Implementation of every round started with judges giving a signal to their players, which in turn picked the card corresponding to their play of choice. Judges where given the score sheet (appendix 3) to record individual players move and score from either of the two players in each round. There where no physical boundaries between the players and they could easily communicate with each other and the judge.

In three sessions, where each session consisted of one class of students, the anchor of altruism where presented, and for the other three sessions, an anchor for altruistic punishment. These anchors were designed to represent different strategies with differing incentives for reciprocal altruism. Anchors where presented alternated by me as a short lecture after game rules and procedures of the game where explained, but before the game actually where played. The anchor of altruism were a short lecture of the importance of niceness for cooperation (appendix 4), while the altruistic punishment anchor put more focus on punishment of defection and its importance for successful cooperation (appendix 5).

The number of rounds to be played where not known to any of the subjects, and there where no time limit. That is, subjects were not able to guess the endpoint of the game, yet 21 rounds were played in all games. In all groups each round where played simultaneously and subjects where given a pay out in cash related to their performance in the game. This is an unusual setup for prisoner’s dilemma games, as players normally either are not able to see or speak to each other (or both), and as judges received a pay out, which was equal to the average of the two players they judged. Using this pay out structure, workload and resources in the form of objective judges were massively reduced. It can also promote better understanding of game rules and its incentives for the participants, but on the other hand, communication within pairs could also trigger a higher than normal rate of cooperation (Sally, 1995; Baillet, 2010). As judges also where given pay outs, mutual cooperation in their pair, would maximize the judges payout (see pay out matrix). Most laboratory experiments in game theory are designed to remove social effects such as kindness, but in open experiments like the present, social effects could influence participants behaviour (Andreoni, 1995).

3.Results

There was no difference in number of cooperative rounds, total score (pay out), number of mutually cooperative rounds or number of mutually defective rounds between treatments (see Table 1).

Table.1: Test statistics

Test	Treatment	Mean Value	SE	N	t value	p value
Cooperative Plays	Altruistic	14.92	0.67	94	0.33	0.74
	Reciprocal	15.24	0.71	94		
Total Score	Altruistic	47.88	1.91	94	0.38	0.71
	Reciprocal	48.89	1.89	94		
Mutual Cooperation	Altruistic	7.79	1.46	74	0.51	0.62
	Reciprocal	8.91	1.68	74		
Mutual Defection	Altruistic	1.42	0.32	74	1.18	0.25
	Reciprocal	0.96	0.22	74		

Table 1: Results from 21 rounds of prisoner's dilemma calculated by unpaired t-test between players anchored with either reciprocal or altruistic anchors. Cooperative plays shows individual player's average cooperative moves. Total score is individual player's average total score in points (see pay out matrix). Mutual cooperation describes number of consecutive rounds were both players cooperated simultaneously. Mutual defection is the number of consecutive rounds were both players defected simultaneously

Additionally, the first defection in each pair and the following moves were analysed specifically with regard to an *a priori* assumption about differences between the two groups in temptation to defect repeatedly. That is, I assumed that temptation of repeated defection would be higher in the group anchored to altruism than to reciprocity (see Tullberg and Tullberg, 1994). In the opening move 16.67 % of the players from the altruistic anchor and 13.04 % from the altruistic punishment anchor defected in their opening move (Yates Chi-squared N = 74, $X^2 = 0.23$, p = 0.63). On average, the first defection within pairs of both anchors came on average after 1.3 rounds (N = 47, SE = 0.85). This result is calculated discarding pairs that cooperated through all rounds of the game. I observed a slight tendency of more cooperation in the group anchored with reciprocal altruism, that is, the frequency of pairs that only cooperated throughout the session in the two groups was 16.7 % for altruism and 26.09 % for altruistic punishment (Yates chi-square N=47, $X^2 =$

1.66, p = 0.19).

I also tested whether players forgave defections differently in the two groups, that is, whether they cooperated after first being defected by their opponent. I found a non-significant difference using unpaired t-test (N = 74, t = 1.1 , p = 0.27, SE = 0.35) as players anchored with altruism played 1.15 (SE = 0.3) cooperative moves after being defected, and players with the altruistic punishment anchor played 0.76 (SE = 0.14) moves. The frequency of defections following players first defection was tested specifically to reveal if free riding strategies differed between groups. I found a significant difference using an unpaired t-test (N = 74, t = 2.57, p = 0.012, SE = 0.25) in the frequency of defections following the first defection by the defecting player. That is, players with an altruistic anchor defected on average 1.05 times (SE = 0.21) in row, while players anchored with altruistic punishment defected on average 0.41 times (SE = 0.1) in a row (see illustration 1).

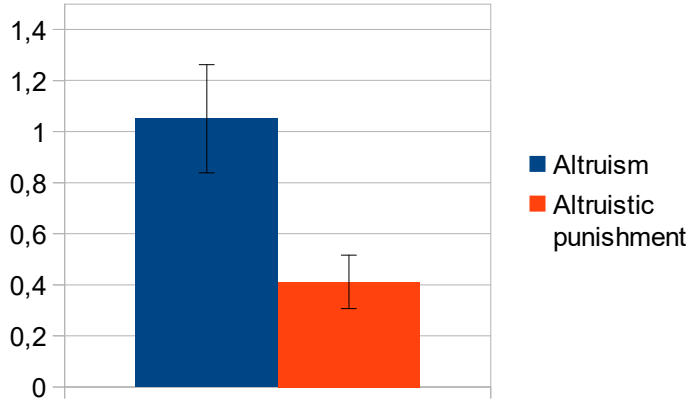


Illustration 1: Mean number of defections following the first defection between the two groups. Error bars indicate standard errors of the mean.

4. Discussion

This study which anchored players of prisoner's dilemma with different cooperative strategies suggest that players do not adopt these strategies to the same extent as in previous research done in this field. I observed no significant difference between the groups in total numbers of cooperative acts, or in the numbers of rounds of mutual cooperation. Although players of both anchors started their games with approximately the same defect-cooperate ratio, the first defection in each game triggered different reactions. My analysis of forgiveness, that is, cooperative plays following the opponents defection were not significant. Yet, my findings revealed that players in groups anchored with altruism defect more frequent succeeding their first defection than players in groups anchored with altruistic punishment. The latter supports that pure altruism makes people

more vulnerable for free-riding (Heinrich and Boyd, 2001), and reciprocal morality are more than capable to motivate humans to behave nice to each other (Tullberg and Tullberg, 1994).

The minority of players in iterate prisoner's dilemma choose to defect in their opening move (Grujic et al., 2012), and this can also be said for players in this study. The average first defection came early (1.3 rounds), and could be contributed by the players learning the incentives of the game (Hilbe and Sigmund, 2010). Also, when comparing simple parameters, such as player's total cooperative or defective rates, I expected to find differences between treatments. My results show, on the other hand, that player's from both groups defect and cooperate in approximately the same fashion. This is also reflected in player's average total scores, which do not differ between groups. These results are in contrast to previous research where participants are more willing to cooperate even in one-shot games when cooperative choices are expected from their partners (Liberman, 2004; Sally, 1995). Even after small manipulations like renaming prisoner's dilemma to "stock market game" or "community game", effects in the predicted directions (i.e., less and more cooperation, respectively) were observed between players in the two treatments (Liberman, 2004; Ellingsen et al., 2012). Although the anchoring process in the present study included specific instructions about how to act, effects are not seen throughout the 21 rounds of the game.

Mutual cooperation in iterated, two players games is a stable strategy (Axelrod and Hamilton, 1981) and could be descriptive of an adopted reciprocal behavior by the players (Stevens et al., 2002). As a stable reciprocal environment were more likely to appear in the group anchored with altruistic punishment, I expected longer cooperative streaks in this group compared to that anchored with altruism. Yet, I found no significant difference in mutual cooperation, nor in 100% cooperating pairs between groups. The altruistic punishing groups had, however, slightly longer cooperative streaks than the pure altruists and more pairs cooperated through the entire length of the game. Social effects, such as kindness (Andreoni, 1995), "the shadow of the future" (Axelrod, 1984) and being observed by their acquaintances in the open structure of the study, might in this case have had a greater effect than the anchors presented. These factors were similar for both groups and would therefore impact players in the same way, even though anchors instructed them otherwise (Sally, 1995; Hilbe and Sigmund, 2010). It is also shown that communication and being watched enhances chances of cooperation (Sally, 1995; Hilbe and Sigmund, 2010). Additionally, being acquaintances might also have had a positive affect on cooperation in this study, as participants performance here may have affected their reputation in their real world daily life more than what would be the case with interactions towards strangers. That is, participants could see,

interact and remember their opponent's choices, choices that would play a part in further interactions after the game itself was finished. Interactions with others that the individual want to keep a good reputation towards, could make the decision not to defect easier than if the game where played anonymously.

Cooperation following an opponent's defection is not a rational choice for any player of prisoner's dilemma (Rapoport and Cammah, 1965; Axelrod, 1984; Milinski and Wedekind, 1998). Still, participants anchored with altruism in this study are encouraged to behave irrationally, while the altruistic punisher's are to never forgive defection. The implicit expectation would therefore be to observe a forgiving attitude from players anchored with altruism, with a close to zero rate of forgiveness from groups of altruistic punishment. In fact, the observed forgiveness within the altruistic group was as expected, as they cooperated on average more than one round after being defected. Achieving even higher rates of cooperation subsequent to defection would have indicated apparent altruistic intentions, but self interest should promote sanctions towards the defecting player to stop free-riding after failing to initiate mutual cooperation through a forgiving move (Fehr and Fischbacher, 2004). The surprising part of this analysis is the high rate of forgiveness in the group of players that was anchored to behave rationally, the altruistic punishing group. As mentioned in the above paragraph, there are several reasons to why players in this setting could cooperate more than expected, but these forgiving choices could also be misplaced altruism or signals of indirect reciprocity. As these games where observed and judged by a third person, altruistic displays to promote their altruistic intentions could potentially benefit the altruistic individual through indirect reciprocity from both the opponent and the observer in future interactions (Tullberg, 2004).

Defective plays following a player's own first defection are significantly higher in the altruistic anchored groups, defectors seem to try to take advantage of the presumed irrational behaviour of their opponents. If players believe that their partners will behave irrationally, the rational play would be to free ride (Andreoni, 1988). The seemingly rational anchor of altruistic punishment, on the other hand, seem to promote a more reciprocal environment, as defections are punished and free riding are deprived. These results are in contrast to previous research (Baillet, 2010; Sally, 1995) where human cooperation is enhanced by increased expectations of partner cooperation. Yet, its been observed that previous moves can effect players mood, hence the term moody conditional players. This describes increased cooperation following their own cooperation and also more defection following their own defection (Grujic et al., 2012; Gutiérrez-Roig et al., 2014). Moody conditional players are observed in most social dilemmas and should therefore be

found in groups within both anchors in this study. That is, free riding is a more likely explanation for the defecting behavior observed in this study, rather than a higher percentile of moody defectors among the presumed altruist's, than among the altruistic punisher's.

The findings in this study show that human teenagers take advantage of others presumed altruistic behavior, if they get told to be purely altruistic. This occurs even though their opponents actually do not adopt that altruistic behavior, and defectors could therefore be better off by being cooperative. The results from the present study suggest that without the sanction methods that are implicit in reciprocal systems, an altruistic system could easily degenerate to a selfish system (Tullberg and Tullberg, 1994). That is, when players expect an environment of altruistic opponents, they find opportunities to take advantage of the presumed altruistic attitude, and when they expect to be punished for selfishness, they show reluctance promote their selfish interests.

5.References

Alcock, J. (2009). Animal behavior (9th ed.). Sinauer associates inc, Massachusetts, USA.

Alexander, R. D. (1974). The evolution of social behavior. Annual review of ecology and systematics, vol. 5, pages 325-383.

Andreoni, J. (1988). Why free ride? : Strategies and learning in public goods experiments. Journal of Public Economics, Elsevier, vol. 37, issue 3, pages 291-304.

Andreoni, J. (1995). Cooperation in public-goods experiments: Kindness or confusion?. The American economic review, vol. 85, issue 4, pages 891-904.

Axelrod, R. (1980). Effective Choice in the Prisoner's Dilemma. The Journal of Conflict Resolution, vol. 24, issue 1, pages 3-25.

Axelrod, R. and Hamilton, W.D. (1981). The evolution of cooperation. Science, vol. 211 , issue 4489, pages 1390 - 1396

Axelrod, R. (1984). The Evolution of Cooperation. Basic Books, New York.

Balliet, D. (2010). Communication and Cooperation in Social Dilemmas: A Meta-Analytic Review. Journal of Conflict Resolution, vol. 54, issue 1, pages 39-57.

Boyd, R. and Richerson, P. J. (1989). The evolution of indirect reciprocity. Social network, Vol. 11, issue 3, pages 213-236.

Chapman, G. B. and Johnson, E. J. (1994). The limits of anchoring. Journal of Behavioral Decision Making, vol. 7, pages 223-242.

Deutsch, M. (1958). Trust and suspicion. Journal of Conflict Resolution, vol. 2, pages 265–279.

Dugatkin, L. A. and Reeve, H. K. (1998). Game Theory and Animal Behaviour. Oxford University Press, Oxford.

Dunbar, R. I. M. (1993). Coevolution of neocortical size, group size and language in humans. Behavioral and Brain Sciences, vol. 16, pages 681-694.

Ellingsen, T., et al. (2012). Social Framing Effects: Preferences or Beliefs?. Games and Economic Behavior. vol. 76, pages 117-130.

Fehr, E. and U. Fischbacher (2004). Social norms and human cooperation. Trends in Cognitive Sciences, vol. 8, issue 4, pages 185-190.

Fehr, E. and Fischbacher, U. (2005). Human altruism–proximate patterns and evolutionary origins. Analyse & Kritik, vol. 27, issue 1, pages 6-47.

Fehr, E. and Gachter, S. (2002). Altruistic punishment in humans. Nature, vol. 415, issue 6868, pages 137-140.

Frank, S. A. (1995). Mutual policing and repression of competition in the evolution of cooperative groups. Nature, vol. 377, pages 520–522.

Gachter, S. and Falk, A. (2002). Reputation and Reciprocity: Consequences for the Labour Relation.

Scandinavian Journal of Economics, vol. 104, issue 1, pages 1-27.

Gutiérrez-Roig, M., et al. (2014). Transition from reciprocal cooperation to persistent behaviour in social dilemmas at the end of adolescence. Nature Communication, vol. 5, article 4362.

Gokhale, C. S. and Traulsen, A. (2010). Evolutionary games in the multiverse. Proceedings of the National Academy of Sciences, vol. 107, issue 12, pages 5500-5504.

Green, D., et al. (1998). Referendum contingent valuation, anchoring, and willingness to pay for public goods. Resource and Energy Economics, vol. 20, issue 2, pages 85-116.

Grujic, J., Eke, B., Cabrales, A., Cuesta, J. A. and Sánchez, A. (2012). Three is a crowd in iterated prisoner's dilemmas: experimental evidence on reciprocal behavior. Scientific Reports, vol. 2, article 638.

Heinrich, J. and Boyd, R. (2001). Why people punish defectors. Journal of Theoretical Biology, vol. 208, pages 79–89.

Hilbe, C. and Sigmund, K. (2010). Incentives and opportunism: from the carrot to the stick. Proceedings of the Royal Society of London B: Biological Sciences, vol. 277, issue 1693, pages 2427-2433.

Jordan, J. J., et al. (2016). Third-party punishment as a costly signal of trustworthiness. Nature, vol. 530, issue 7591, pages 473-476.

Kiyonari, T. and Barclay, P. (2008). Cooperation in social dilemmas: Free riding may be thwarted by second-order reward rather than by punishment. Journal of Personality and Social Psychology, Vol. 95, issue 4, pages 826-842.

Liberman, V., et al. (2004). The Name of the Game: Predictive Power of Reputations versus Situational Labels in Determining Prisoner's Dilemma Game Moves. Personality and Social Psychology Bulletin, vol. 30, issue 9, pages 1175-1185.

Maynard Smith, J. (1982). Evolution and the theory of games. Cambridge, Cambridge University

Press.

Milinski, M. and Wedekind, C. (1998). Working memory constrains human cooperation in the Prisoner's Dilemma. PNAS**,** vol. 95, issue 23, pages 13755-13758.

Milinski, M., et al. (2002). Reputation helps solve the tragedy of the commons. Nature, vol. 415, issue 6870, pages 424-426.

Nowak, M. A. and Sigmund, K. (1992). Tit for tat in heterogeneous populations. Nature, vol. 355, issue 6357, pages 250-253.

Nowak, M. A and Sigmund, K. (2005). Evolution of indirect reciprocity. Nature, Vol. 437, issue 7063, pages 1291-1298.

Oliver, P. (1980). Rewards and punishments as selective incentives for collective action: theoretical investigations. American Journal of Sociology, vol. 85, issue 6, pages 1356–1375.

Phan, K. L., Sripada, C. S., Angstadt, M., McCabe, K. (2010). Reputation for reciprocity engages the brain reward center, PNAS. vol. 107, issue 29, pages 13099-13104.

Rapoport, A. and Chammah, A. M. (1965). Prisoner's dilemma : a study in conflict and cooperation. Ann Arbor, University of Michigan Press.

Rapoport, A. and Guyer, M. (1966). A taxonomy of 2 x 2 games. General systems, Vol. 11, pages 203-214

Ratnieks, F. (1988). Reproductive Harmony via Mutual Policing by Workers in Eusocial Hymenoptera. The American Naturalist, vol. 132, issue 2, pages 217–236.

Sally, D. (1995). Conversation and Cooperation in Social Dilemmas:A Meta-Analysis of Experiments from 1958 to 1992. Rationality and Society, vol. 7, issue 1, pages 58-92.

Stephens, D. W., et al. (2002). Discounting and Reciprocity in an Iterated Prisoner's Dilemma. Science, vol. 298, issue 5601, pages 2216-2218.

Trivers, R. L. (1971). The evolution of reciprocal altruism. Quarterly Review of Biology, vol. 46, issue 1, pages 35–57.

Tullberg, J. and Tullberg, B. (1994). Naturlig etik: En uppgjörelse med altruismen. Lykeion, Stockholm

Tullberg, J. (2004). On Indirect Reciprocity: The Distinction between Reciprocity and Altruism, and a Comment on Suicide Terrorism. The American Journal of Economics and Sociology, vol. 63,issue 5, pages 1193–1212.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, vol. 185, issue 4157, pages 1124–1130.

Van Lange, P. A. M., Klapwijk, A. and Van Munster, L. M. (2011). How the Shadow of the Future might Promote Cooperation. Group Processes & Intergroup Relations, vol. 14, issue 6, pages 857–70.

Wilkinson, G. S. (1984). Reciprocal food sharing in the vampire bat. Nature, vol. 308, issue 5955, pages 181-184.

6.Appendix

6.1 Payout matrix

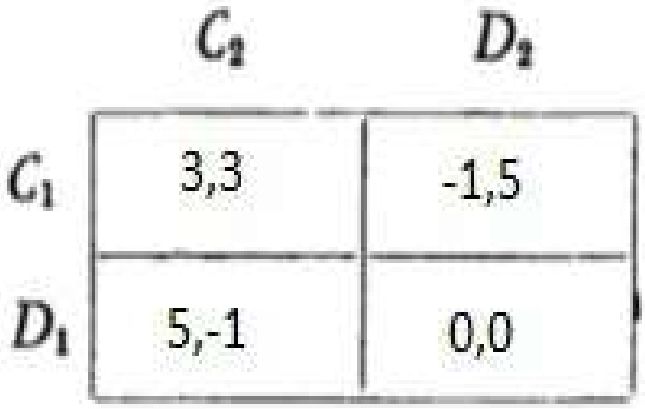


Illustration 2: Payout matrix two player prisoners dilemma with points as assigned in this experiment

6.2 Score sheet

Player 1			**Player 2**	
Round	Action	Points	Action	Points
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
Total points player 1:			Total points player 2:	

6.3 Seating diagram

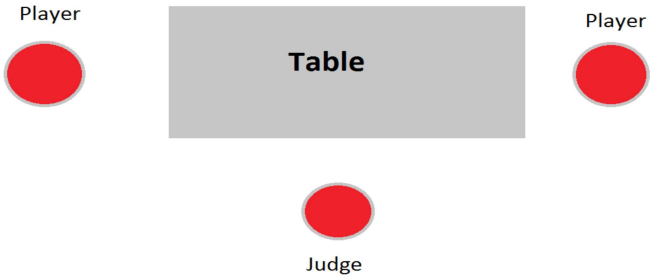


Illustration 3: Seating diagram for two players and judge.

6.4 Anchor for altruism

Altruism (Translated from Norwegian)
Before we begin playing the game, I wish to remind you of the values that has been important for us humans since our dawn.

Humans are social animals that live together in large groups. For groups with a large spectre of human interests and personalities to function, a moral code should be present to guide us in interaction with our fellow citizens. Without such a base, groups would pulverize in to individual's own interests and therefore stop functioning. Each individual's selfish motifs would rule and the synergy effects with cooperation would vanish. Members of the rock band “Foo Fighters” could for example never establish such a popularity one by one as they did together. Our society are built on moral building blocks that yields cooperation through laws and rules for behavior - Behavior that promotes synergy and companionships profit that exceeds the profit of individuals.

These values are built on being nice to each other, but also **forgive** those who don't show the same kindness towards you. Why should one **forgive** those individuals who act selfishly? Kindness develops **kindness**, and by being generous and react to defection by **sacrifice** would cooperation

establish easier among humans. Kindness to however come at a cost, as signalling ones kindness could lead to a danger of exploitation by nasty individuals, but by being **genuinely kind**, other will adapt that **kind attitude**. It is also important to remember that life goes over an iterate number of interactions similar to those of prisoner's dilemma - one can never know how many times in a life you would have a interaction with the same person. Its important that one remember that **kindness develops kindness**, and in the long run it's better for **everyone to be kind**.

Summarised, to achieve **genuine kindness** one must be **forgiving**. To handle defections is not necessary easy, but **forgiveness** of defection is important to help achieving a higher percentile of **kindness and forgiveness** by the counterpart. One should therefore be **unlimited kind** and have the **best intentions** in every situation and **forgive** everyone that does not show the same grace. If one could achieve such an attitude, **kindness** in others would increase and give gains not only for the individual, but for all. Similar to the example of “Foo fighters”.

6.5. Anchor for altruistic punishment

Altruistic punishment (Translated from Norwegian)
Before we begin playing the game, I wish to remind you of the values that has been important for us humans since our dawn.

Humans are social animals that live together in large groups. For groups with a large spectre of human interests and personalities to function, a moral code should be present to guide us in interaction with our fellow citizens. Without a such a base, groups would pulverize in to individual's own interests and therefore stop functioning. Each individual's selfish motifs would rule and the synergy effects with cooperation would vanish. Members of the rock band “Foo Fighters” could for example never establish such a popularity one by one as they did together. Our society are built on moral building blocks that yields cooperation through laws and rules for behavior - Behavior that promotes synergy and companionships profit that exceeds the profit of individuals.

These values are built on being nice to each other, but also **punish** those who don't show the same kindness towards you. Why should one **punish** those individuals who act selfishly? Kindness develops **exploitation**, and by being generous and react to defection by **punishment** would cooperation establish easier among humans **of same values**. Kindness to however come at a cost, as

signalling ones kindness could lead to a danger of exploitation by nasty individuals, but by being **watchful**, other will adapt that **attitude to decrease the advantage of defectors**. It is also important to remember that life goes over an iterate number of interactions similar to those of prisoner's dilemma - one can never know how many times in a life you would have a interaction with the same person. It's important that one remember that kindness **and punishment for defectors would be the best strategy**, and in the long run it's better for **every individual.**

Summarised, to achieve **reciprocal behavior** one must be **kind, but also punish defectors**. To handle defections is not necessary easy, but **punishment** of defection is important to help achieving a higher percentile of **cooperation** and **punishment** by the counterpart. One should therefore be **careful,** have **good** intentions in every situation and **not** forgive everyone that does not show the same grace. If one could achieve such an attitude, **cooperativeness** in others would increase and give gains not only for the individual, but for all. Similar to the example of “Foo fighters”.