# Identification of Altruism among Team Members: Empirical Evidence from the Classroom and Laboratory

Griffiths, Barry

University of Central Florida

7 May 2010

# Identification of Altruism among Team Members: Empirical Evidence from the Classroom and Laboratory

by

Barry J. Griffiths

**Abstract**

It is difficult to identify acts that are purely altruistic, and do not have some level of egoism or self-interest involved. By considering situations where team members seemingly have nothing to gain by the way they distribute points to others with regard to peer evaluation, and where their acts of giving are anonymous, can never be verified, and cannot guarantee reciprocity, we see that there is still significant evidence of altruism, which we then confirm using laboratory experiments.

Draft: February 2015

Keywords: Altruism, individual behavior, group behavior, laboratory experiments, reciprocity

**Barry J. Griffiths**: Department of Mathematics, College of Sciences, University of Central Florida, Orlando, FL 32816; barry.griffiths@ucf.edu. I appreciate the comments, advice and/or help we have received from Glenn W. Harrison, E. Elisabet Rutström, and Jeffrey Allen.

Much of the literature concerned with the identification of altruism has had to contend with a counterargument that the actions documented are not in fact altruistic, and stem from the natural instinct of egoism and self-advancement. Batson and Shaw [1991] contend that while the subject of the existence of altruism has been debated since Aristotle, the majority view among current social scientists is that "We are, at heart, purely egoistic, that we care for others only to the extent that their welfare affects ours." The challenge for those who believe that pure altruism exists is therefore to conduct experiments where subjects act without any regard for their own self-interest.

This study starts with field data taken from two courses taught in the Marketing Department at the University of Central Florida, and considers the assessment results from students towards their semester-long team members, having been informed in advance that the component would account for 20% of each student's final grade in the class.

The evidence indicates that team members showed a tendency to apportion points uniformly, even when there were obvious disparities in their individual level of knowledge with regards to the subject material. This suggests that when placed in a group setting, individuals are more likely to overlook known deficiencies, and award points in a manner which tries to ensure the welfare of each of their team members. In order to try and verify this hypothesis an experiment was conducted in a laboratory setting, and designed in such a way as to increase the internal validity of the findings, and overcome the usual confounds with regards to altruism.

# 1. IDENTIFYING ALTRUISM

## *1.1 Historical and Modern Arguments Regarding Altruism*

While the term "altruism" was not used until 1851 [Comte, 1851], it is a concept discussed by philosophers for centuries, dating back to Plato and Aristotle, and developed throughout history by famous figures such as St. Thomas Aquinas, Hobbes, Kant, Hume, and Nietzsche[1]. While the majority took the view that acts of apparent altruism are often motivated by subtle benefits to the perpetrator, this has been challenged in more recent times by economists, biologists, and psychologists to the extent that there are now many who believe that altruism is an inherent quality of human nature.

During the last fifty years, research into altruism looked less at philosophical arguments, and more at how it relates to evolution, thereby advancing the work of Charles Darwin and Herbert Spencer. In "The Selfish Gene," Richard Dawkins [1976] presented a genetic view of evolution, and contended that while genes act "selfishly", they are not driven by any motives or will – merely that their effects can be accurately described *as if* they do. The contention is that the genes that get passed on are the ones whose consequences serve their own implicit interests (to continue being replicated), not necessarily those of the organism, much less any larger level. This view explains altruism at the individual level in nature, especially in kin relationships. When an animal sacrifices its own life to protect the lives of kin, it is acting in the interest of its own genes.

The notion of kin altruism (though he did not use the term) was formally developed by Hamilton [1964], who gave the hypothesis (known as Hamilton's Rule) that altruistic behavior occurs when $br > c$, where b is the benefit, r is the relatedness of the individuals concerned and c

[1] Comte coined the French word *altruisme* from the Italian *altrui,* meaning "of others". The *Oxford English Dictionary (Second Edition)* states that altruism is "devotion to the welfare of others, regard for others, as a principle of action: opposed to egoism or selfishness."

is the cost associated with the actions. For humans, the value of r is 1 for identical twins, 0.5 for parents and their offspring, as well as full siblings, 0.25 for grandparents and grandchildren, as well as half siblings etc. This theoretical view of kin altruism is consistent with the mathematical foundations of evolutionary theory proposed in the 1930's by the biologist J.B.S. Haldane[2], who joked at the time that "I would lay down my life for two brothers or eight cousins." [Connolly & Martlew, 1991]

John Maynard Smith [1988], a leading expert in the field of evolutionary biology who studied under Haldane, considered these theories of kin altruism to be false, and that within the group there would always be some level of free-riding from individuals seeking to benefit from the altruistic acts of others while not incurring their fair share of the cost. He looked at altruism from a game theoretic point of view, and felt that unlike with humans who have the capacity to rationally consider strategies and their consequences, that organisms automatically do what is best to propagate their population. He dismissed the notion of "group selection", whereby group survival determines individual behavior, considering it to be an inherently weak evolutionary force, and hence unlikely to promote interesting altruistic behavior.

Regardless of the conclusions drawn from the study of kin altruism, the fact that apparent acts of altruism were observed among non-related species dictated a broader approach to the overall theory. Trivers [1971] developed the notion of reciprocal altruism, which is not restricted to kin, whereby a person or animal will act kindly towards others with the expectation that the act will be reciprocated in the future (an underlying assumption being that there will be further interaction, and that it is not a "one-shot game"). Although this area has been studied in depth, many social scientists believe that everything they need to know about reciprocity is summarized

[2] Haldane asserted that on average a person shares 50% of their genes with siblings and each parent, 25% with aunts and uncles, and hence 12.5% with cousins.

by the Tit-for-Tat strategy in the Prisoner's Dilemma game, where a Nash equilibrium is reached if both players decide to adopt it.

The popularity of Tit-for-Tat derives from Axelrod's 1984 Olympiad (and subsequent 1986 ecological tournament), where strategies were matched against each other in the indefinitely repeated Prisoner's Dilemma game. Although Tit-for-Tat, entered by Rapoport, performed very well due to being nice, retaliatory, forgiving, and clear, it was then proposed as a suitable paradigm for human behavior across the board. However, although the results are cited by approximately 300 research papers each year according to Axelrod, the success of the strategy came to be somewhat embellished, especially by those determined to believe that humans are essentially nice, and depends largely on the initial population of entries. Binmore [1998, 2006] described the data as being "woefully inadequate", and that "...[Prisoner's Dilemma] represents a situation in which the dice are as loaded against the emergence of cooperation as they could possibly be." Rapoport himself [1991] wrote that "The reason Tit-for-Tat won both contests is because the "more aggressive" or "smarter" strategies beat each other."

While the investigations into reciprocal altruism have continued, the focus of the experiments has broadened to consider not just the biological and evolutionary aspects of altruism, but also to provide scientific rigor to the psychological aspects of why people often act in an altruistic manner. However the difficulty for researchers has been convincing skeptics that the apparent acts of altruism are not merely acts of egoism in disguise. A discussion on how to experimentally discern between altruism and egoism was provide by Batson & Shaw [1991], who were concerned about the distinction between altruism and "pseudo-altruism" (where engaging in altruistic deeds lead to an obvious advantage) . They concluded that the opinions of those who always believe that self-interest is involved "…must give way to a pluralistic explanation that includes altruism as well as egoism."

Batson & Shaw's article was critiqued by 15 commentators, the general argument being between those who agreed that humans are naturally altruistic, and those who maintained the historical view that such acts are always accompanied by "…the stubborn presence of less salubrious motives." [Scott & Seglow, 2007]

### *1.2 Economic Implications of Altruism*

Although acts of altruism are generally viewed in a positive light, considerable research and debate has focused on whether they can develop into "too much of a good thing" when taken to extremes, in particular when incorporated into government policy. Coate [1995] argued that "altruism provides an efficiency rationale for public provision of insurance to the poor…[and that] when the government makes unconditional transfers, the poor may have an incentive not to buy insurance and to rely on private charity to bail them out in the event of loss." This has been seen most clearly with regard to natural disasters such as floods and hurricanes, where compensation provided by the government and charitable organizations has mitigated the need for individuals to insure themselves against the occurrence of such events. Such exploitation of good intent illustrates the ideas of Buchanan [1975], who introduced the "Samaritan's Dilemma", which claimed that the availability of assistance leads more people to demand it, thus increasing the number of people in need. Buchanan concluded that "We may be simply too compassionate for our own well-being or for that of an orderly or productive society." [3]

[3] The ideas of Buchanan have been seized upon by right wing commentators and politicians, especially in the U.S. as the basis for eliminating altruism from government policy. Stone [2008] dismissed such hard-heartedness, and claimed that Buchanan's work merely "…reveals how ideologues can use specious science to undermine human hope and morality."

Stark [1989, 1995] offered a less dismissive view, while cautioning that although altruism as an attribute is good, it does not necessarily result in improved outcomes, due to the increased possibility of exploitation and strategic behavior. He came to the conclusion that "…while an economy with substantial altruism will be Pareto superior to an economy with no altruism, an economy with a little altruism may be inferior to an economy with no altruism at all."

Such arguments have been much discussed in recent times given the assistance given in 2009 by the US government to homeowners in danger of foreclosure. Altruism has been viewed as both the cause of the problem, due to the decreased lending standards encouraged by the government to promote home ownership among minorities [Salsman, 2009], and also the solution[4] to avoid a more serious economic recession.

### *1.3 Evidence of Altruism in Behavioral Experiments*

In his book *The Gift Relationship* [1970], Titmuss develops the "crowding out" theory, whereby people will donate money (or in particular blood) without any incentive if they believe it is for the common good, but if a reward is offered, then their impulse to act will be reduced. Solow [1971] offered a skeptical view due to the lack of analytic rigor, stating that "Titmuss makes a very weak case for this belief (that altruism is more pervasive than economists care to acknowledge)" and concluded by adding that "[Titmuss] seems to believe that ordinary people ought to be happy to have many decisions made for them by professional experts who will, fortunately, often turn out to be moderately well-born Englishmen." However the crowding out theory was vindicated with regard to blood donation when an experiment conducted by Upton [1974] found that a blood drive in Kansas City and Denver attracted 93% of those asked to denote voluntarily, but only 65% of those offered $10 to participate.

[4] The Homeowner Affordability and Stability Plan, announced in February 2009, involved a $75 billion federally funded program to modify the mortgage terms of borrowers at risk of foreclosure.

One of the seminal behavioral experiments looking at altruism was carried out by Güth, Schmittberger, and Schwarze [1982], who investigated behavior among subjects playing the Ultimatum Game, whereby Player A is initially given $10 and Player B is initially given nothing. Player A offers a share of the money to Player B, but a rejected offer means that both players end up with nothing. While theory would indicate that even a $1 offer should be accepted, in practice the median offer is usually more than $4 and responders reject offers of $2 half the time. Camerer [2003] explains this apparent discrepancy by stating that "Generous offers could come about because Proposers are fair-minded or because Proposers are afraid of having low offers rejected (or both)."

This distinction is clarified by the Dictator Game, which removes the ability to reject an offer. In the original experiment run by Kahneman, Knetsch, and Thaler [1986] dictators were given the option of splitting $20 evenly or splitting the money ($18, $2). 76% chose the even split, indicating a fair-minded approach. A later experiment by Forsythe, Horowitz, Savin, and Sefton [1994] found that such generosity was reduced when real money was involved, with only 21% dictating an even split of a $10 pot. However 79% offered at least the minimal amount of $1 (the mean being $2.33), indicating a reluctance to leave an unknown partner with nothing.[5]

Given all the experimental evidence, it became clear that the traditional economic models that rendered altruistic acts as being irrational required modification. The derivation of new models was needed to explain the presence of altruism, consistent with theories of equilibrium analysis, which could be translated into policy. However, as Sugden [1984] noted, "It is remarkably difficult to produce a satisfactory theory to explain how public goods come to be

[5] The primary distinction between the two experiments is that in the Kahneman et al. study the probability that a subject would actually be paid was only 0.05, whereas in the Forsythe et al. study participants received the full monetary amount from the result of the game, in addition to a participation fee. Forsythe et al. found statistical evidence to reject the hypothesis that the distributions of proposals are the same with or without pay.

supplied through voluntary activity when many individuals are involved. The problem, of course, is the incentive for each individual to take a free ride."

Becker [1974] proposed a model of pure altruism, which assumed utility maximization, and that individuals take as given the contributions of others. Although the welfare of others appeared as a positive argument in an individual's utility function, the model lacked predictive power, and was advanced by Andreoni [1990], who generalized the work of Becker to include the notion of "impure altruism", where an individual's utility depends on the welfare of others as well as the amount contributed (known as the Warm Glow Theory).

However Sugden [1982] observed that in pure altruism models, players with a concave utility function will reduce their contribution when the other members of their group donate more, stating that "If many people are contributing to a public good, it turns out that any increase on one person's contribution is almost completely offset by decreases in other people's contributions." Citing the lifeboat service and the blood transfusion service in the U.K. as counterexamples, he went on to propose a different model (see Sugden [1984]), based on the "principle of reciprocity," whereby individuals choose a level of contribution based on what they would prefer every other member of the group to make.[6] If we assume this morally binding constraint, then the model does not contain the supposition of free-riding which exists in the models of Becker and Andreoni, and instead leads to a positive correlation between a donor's gift and the amount given by the other members of the group.

---

[6] Sugden formally describes the principle of reciprocity as follows: "Let *G* be *any* group of people of which *i* is a member. Suppose that every member of *G* except *i* is making an effort of at least $\xi$ in the production of some public good. Then let *i* choose the level of effort that he would most prefer that every member of *G* should make. If this most preferred level of effort is not less than $\xi$, then *i* is under an obligation to the members of *G* to make an effort of at least $\xi$."

Many other models have since incorporated a reciprocity component, with the theory of "strong reciprocity" (see Gintis [2000], Fehr and Gächter [2000a,b]) being a significant area of study. Here participants are not only predisposed to cooperate with others, but will punish non-cooperators. This has led to the explanation of many behavioral observations that cannot be justified by self-interest, kinship or reciprocal altruism.

## 2. COLLECTION OF DATA FROM CLASSROOM

Data was collected from two undergraduate (junior and senior level) classes taught by Dr. Jeffrey Allen in the Marketing Department at the University of Central Florida, with 95 students in total. The classes were split into groups of five or six students, with graded work including quizzes, case studies, and a group project. Each (multiple choice) quiz was taken three times, with the questions repeated. On the first occasion students worked on an individual basis, secondly in consultation with other group members, and thirdly as a group with notes and the textbook being permitted. In every case, as one would expect, the scores increased with each attempt. Results for each student were then quickly calculated by a Scantron machine and returned by the end of the lecture, with the scores being tallied by a member of each group in full view of the others. Consequently each member had a precise empirical knowledge of the ability of their teammates, as well as the impressions formed during the group discussions. These were augmented by the lengthy projects given to each group, which required meeting outside of the lectures, and caused a further reduction in the social distance[7] between team members.

At the end of the semester students were asked to distribute an average of 10 points per team member (so 50 points in a group of six – students not being able to award points to

[7] According to Hoffman et al. [1996], "Social distance can be defined as the degree of reciprocity that subjects believe exist within a social interaction." In experiments based on the Dictator game they found that participants became less generous as social distance increased.

themselves), with the stipulation that the points could not be uniformly distributed (see the third instruction in Figure 1). The points were awarded privately, and the results not released. The evaluation of each student by their team members contributed 20% towards their final grade in the class.

**Figure 1 – Peer Evaluation Instructions**

*PEER EVALUATION*[8]

*NAME ____________________* *GROUP # _____*

*Please try to assign scores that reflect how you really felt about the extent to which the other members of your group contributed to your learning and/or your group's performance. This will be your only opportunity to reward the members of your group who actually worked hard on your behalf. If you give everyone pretty much the same score you will be hurting those who did the most and helping those who did the least.*

*Instructions: In the space below please rate each of the other members of your group. Each member's peer evaluation score will be the average of the points they receive from the other members of the group. (As a result, the average peer evaluation score for all the groups will be identical.) To complete the evaluation you should:*

*1. List the name of each of the members of your group in the alphabetical order of their last names.*

*2. Assign an average of ten points to the other members of your group. (Thus, for example, you should assign a total of 50 points in a six member group; 60 points in a seven member group, etc.)*

*3. Differentiate some in your ratings, e.g., you must give at least one score of 11 or higher and one score of 9 or lower.*

| GROUP MEMBERS | SCORES | GROUP MEMBERS | SCORES |
|---|---|---|---|
| 1) ____________________ | ____ | 4) ____________________ | ____ |
| 2) ____________________ | ____ | 5) ____________________ | ____ |
| 3) ____________________ | ____ | 6) ____________________ | ____ |

*Additional Feedback: In the space below would you also briefly describe your reasons for your highest and lowest ratings. These comments – but not the information about who provided them – will be used to provide feedback to students who would like to receive it.*

*1. Reason(s) for your highest rating(s). (Use back if necessary.)*

*2. Reason(s) for your lowest rating(s). (Use back if necessary.)*

[8] The original instructions were not italicized.

# 3. COLLECTION OF DATA FROM LABORATORY

## *3.1 Experimental Design*

Given that peer assessment is not typically included in grade determination, it was necessary to design an experiment which would either validate or reject the results of the field data. This was done in such a way as to balance the need to explain the effects of any confounds from the original data (discussed in Section 3.2), while maximizing the internal and external validity.

The three experiments took place in the Behavioral Laboratory on the campus of the University of Central Florida. Each experiment involved 25 undergraduate students. The first group acted as the control, the second group was divided into numbered teams of 5 participants randomly formed by the experimenter, and the third group was divided into numbered teams of 5 participants formed in advance of the experiment from a senior level class on Game Theory taught in the previous semester.

### Experiment 1: No Familiarity Among Team Members (Control Group)

This experiment consisted of five parts:

i) Before the main experiment began, participants were asked to play two short games using the Veconlab software. Risk attitudes were measured using the game created by Holt & Laury [2001], and each participant's level of altruism was measured using the version of the Dictator game designed by Forsythe et al. [1994].

ii) The first part of the main experiment involved all participants answering a series of 30 general knowledge questions taken from the Raven Test, which measures abstract reasoning ability by asking participants to identify the missing segment to complete a larger pattern. Each correct answer was worth $0.50.

iii) Once all participants had completed the Raven Test, they were asked some simple demographic questions.

iv) The second part of the main experiment involved participants being given the following instructions:

*You have been randomly grouped with four other people in the room, whose identities will remain anonymous. Their scores from the 30 question test you took earlier were (the numbers quoted here being for illustrative purposes only):*

*Group Member 1 scored 22/30* *Group Member 2 scored 13/30*

*Group Member 3 scored 20/30* *Group Member 4 scored 25/30*

*Along with your own score of 15/30, this gives an average score for your group of 19/30. Adding these scores means that 15 + 19 = 34 x $0.50 = $17 will be added to your total earnings.*

v) For the final part of the experiment participants were invited to a separate room, where they were given the following instructions.

*You now have $20 to distribute among your team members. You can distribute the money any way you want based on the experiments that you have completed. None can be kept for yourself. The results will be kept secret. Once everyone has finished, you will be told to leave the testing lab and collect the money you have earned from this experiment.*

Experiment 2: Low Levels of Familiarity Among team Members

This experiment consisted of five parts, the first three parts being identical to those in Experiment 1. This time however, after completing the demographic questionnaire, participants were given the following instructions:

*You will now take another 30 question test. The five participants in your row will collaborate, with one of you answering on behalf of the group. Be sure to familiarize yourself with the names of the other members of your group.[9] You will be given 20 minutes to complete the test. $0.50 will be given to each group member for every correct answer.*

Once this was completed, the scores were tallied, and the following information passed back to each participant (the numbers quoted here being for illustrative purposes only):

*In the first part of the experiment your score was 21/30.*

*In the second part of the experiment, the score of your group was 25/30.*

*Adding, this means that 21 + 25 = 46 x $0.50 = $23 will be added to your total earnings.*

Participants were then asked to return to their original computer, before being individually invited to a separate room, where they were given the following instructions.

*You now have $20 to distribute among your team members[10].*

*You can distribute the money any way you want based on the experiments that you have completed. None can be kept for yourself. The results will be kept secret. Once everyone has finished, you will be told to leave the testing lab and collect the money you have earned from this experiment.*

---

[9] Name tags were worn by all participants during the second and third experiment.

[10] It is possible to make the case that a $20 amount implicitly encouraged participants to award a uniform distribution of $5 to each of their four team members, given the simplicity of the arithmetic, but this is mitigated by having sophisticated participants.

After a short delay in order to allow amounts to be tallied, participants were paid before leaving the lab.

Experiment 3: High Levels of Familiarity Among Team Members

While waiting to enter the laboratory, participants were asked to form groups of five based on familiarity. Each group then entered the laboratory and was instructed to sit in the same row. The same procedures as in Experiment 2 were then followed.

### *3.2 Addressing Potential Confounds*

It is important to discuss the methodological confounds that may diminish or embellish the altruistic feelings reflected by the field data (shown in Section 4.1), and how they were addressed in the design of the laboratory experiment. Two potential confounds that generally arise in such studies are discussed in detail by Harrison & Johnson [2006]. The first is that “the act of giving is typically correlated with extracting surplus from the experimenter.” Therefore what appears to be altruistic behavior is merely a desire to ensure that no money is left on the table. The second is that “preferences with respect to altruism are likely to be heterogeneous across subjects.” This can be a problem when the sample size is insufficient to overcome the behavioral differences that occur solely based on its composition.

Given that a fixed number of points were to be distributed by each student, the first of these confounds is not a concern. However the second is relevant given that during the time spent together, the members of each group will likely have developed feelings and opinions that are not related to academic ability, which may have affected the distribution of points. As concluded by Andreoni and Miller [2002], “…many things other than the final allocation…are likely to matter to subjects.” The experiment therefore begins with the Dictator Game, which gives an *a priori* measure of altruism.

Cherry et al. [2002] posited that subjects will act in a more altruistic manner when the giving does not involve any sacrifice, and that "Strategic concerns − not fairness − appear to be the motivation for other- regarding behavior when people bargain over earned wealth." While the laboratory experiment did not involve the distribution of earned money, the fact that the total amount accumulated by each participant was dependent on the performance of other group members gave an incentive to punish those who did not contribute a reasonable share and reward those who increased the amount received.

The unusual nature of the exercise that led to the field data – students being asked to evaluate their peers, with the scores being used in the determination of grades – may have caused it to be dismissed as an exercise not worthy of their extended consideration.[11] However the simplicity of the experiments tried to ensure that confusion was not an issue in the laboratory.

In Section 4.1 it is noted that some evidence of collusion was observed when analyzing the field data. Given that the students knew in advance that peer assessment would be a part of their grade, the possibility existed for "cheap talk", which occurs (see Harrison [2006], p. 137) when "…participants may send signals to each other costlessly, and where those signals do not change the payoffs from adopting alternative strategies." Furthermore, the instructions given to students (shown in Section 2) could be interpreted as "cheap talk" script on the part of the instructor if one believes that they encouraged the students to distribute the points in an 11, 10,…, 10, 9 manner. However this was not a factor in the laboratory experiments, given that participants did not know in advance that they would be evaluating their team members, and did so individually and anonymously. The instructions given by the experimenter allowed

[11] Instructions 2 and 3 on the Peer Evaluation form shown in Section 2 could be viewed as a suggestion to try and award points in a manner which deviates from uniformity as little as possible. Although there were 58 instances of this during the field experiment (from 95 total participants), it is likely due to collusion on the part of group members (discussed in Section 4.1). The instructions given during the laboratory experiment clearly informed participants that there were no restrictions on how money could be distributed.

participants to distribute the $20 among their team members without restriction, and without providing examples.

The experiment was also designed to overcome the "checklist" of factors that can lead to easily confounded data, of the type identified by Grether and Plott [1979]. By paying participants (generously) in cash, the issues of misspecified incentives and indifference are addressed, and while gathering participants from the same university campus is far from ideal in terms of generalizing the results obtained to more general populations, each participant's status as a university student minimizes the potential confound that could arise from having unsophisticated subjects.

The possibility of a wealth effect in the laboratory data, i.e. the results being affected by the amounts accumulated by participants, is minimal given that each participant was asked to allocate the same amount of money among members of their group. It should also be noted that participants could not be completely sure of their absolute or relative amount accumulated given that it depended on how much they received from their group.

Finally, issues involving strategic responses, as well as the problems associated with reciprocity and retaliation discussed by Fehr and Gächter [2000a,b] are not relevant here given that the monetary amounts awarded by each participant were not known to other group members.

# 4. RESULTS

## *4.1 Results from the Classroom*

The field data was obtained from the Excel spreadsheets created by Dr. Jeffrey Allen, which included the individual scores associated with each quiz and the average score each student received from the members of their group. A visual inspection of the graphical data indicates that the scores from the individual results of the graded work resulted in a typical distribution (see Tables 1 and 2 below). This is corroborated statistically by the Shapiro-Wilks test for normality.[12]

**Table 1 – Distribution of Quiz Totals from MAR 3613**

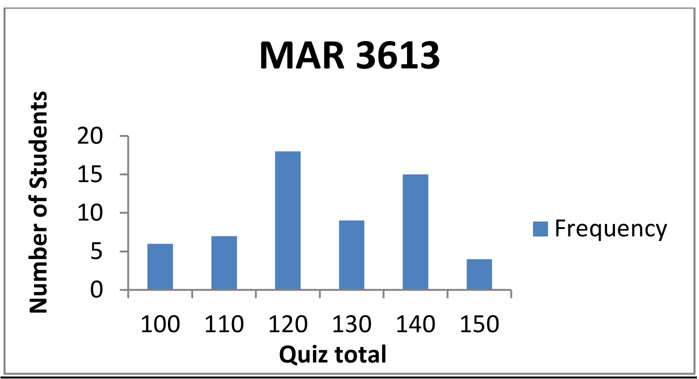
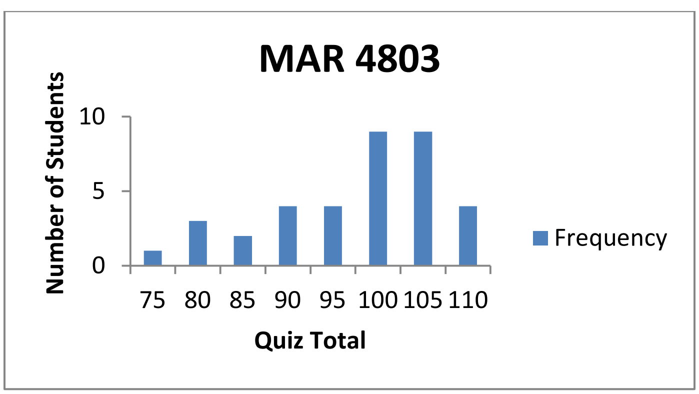



**Table 2 – Distribution of Quiz Totals from MAR 4803**




[12] The Shapiro-Wilks test, when applied to the quiz scores for MAR 3613 and MAR 4803 gave respective test statistics of W = 0.978 (p=0.367) and W = 0.959 (p=0.264), indicating that we cannot dismiss the hypothesis that the scores are normally distributed.

In contrast, the peer evaluation numbers showed an overwhelming propensity to ignore the academic merit of other team members and allocate points in as uniform a manner allowed under the rules, with over 90% of the students receiving a score between 9 and 11.

There was some evidence of collusion among team members to ensure that each member received an average of 10 points. On five occasions (from the nineteen groups in the two combined classes) each member of the team distributed their points in an 11, 10,…, 10, 9 manner, with the highest and lowest scores alternating in such a way that each person received them on one occasion, thus ensuring a uniform distribution overall.[13]

The hypothesis that students did not correlate their discretionary points with individual performance was then tested statistically (see Tables 3 and 4)[14], using a Tobit model to show that the scores awarded by the students could not be predicted with any level of statistical significance by the level of performance on the individual quizzes.

**Table 3 – Tobit Model Results from MAR 3613 (59 students)[15]**

| Variable | Coefficient | Standard Error | t | P>\|t\| | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| (Constant) | 8.11 | 0.89 | 9.13 | 0.000 | 6.33 | 9.89 |
| Quiz | 0.02 | 0.01 | 1.16 | 0.246 | 0.00 | 0.03 |

[13] In both classes the descriptive statistics indicated that the scores awarded by the students did not have the same level of dispersion as those from the individual work. The coefficient of variation (the ratio of the standard deviation and the mean) for the quiz scores was 0.213 (MAR 3613) and 0.182 (MAR 4803), whereas for the peer evaluations the respective coefficients were 0.077 and 0.094, indicating reluctance to award extreme scores.
[14] There is an implicit assumption here that students were awarding points based on productivity rather than effort, as the only variable in the model is the raw quiz score. However the instructions given, "Please try to assign scores that reflect how you really felt about the extent to which the other members of your group contributed to your learning and/or your group's performance", could be interpreted as asking students to consider effort also, which can lead to different results (see Rutström & Williams [2000], where the number of solutions found and the number of moves made during a 30 minute experiment involving the Tower of Hanoi puzzle was used to distinguish between the two effects). The comments that were provided by students to support their allocations show some evidence of this, with penalties for lack of attendance and lack of interest being more prevalent than penalties for lack of knowledge.
[15] A two-sided Tobit model was used given that the quiz scores were constrained to [0, 45].

The correlation coefficient (0.073) indicates a very weak correlation between the scores awarded to each student by their peers and their performance on the individual quizzes. Furthermore the level of significance (p = 0.585) would not lead one to conclude that the two are correlated.

**Table 4 – Tobit Model Results from MAR 4803 (36 students)**

| Variable | Coefficient | Standard Error | t | P>\|t\| | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| (Constant) | 8.53 | 1.42 | 6.01 | 0.000 | 5.65 | 11.41 |
| Quiz | 0.02 | 0.01 | 1.04 | 0.305 | -0.01 | 0.05 |

Again, the level of correlation (0.171) is very weak between the peer evaluation scores and those gained on the quizzes. The significance level (p = 0.305) would lead one to dismiss any hypothesis that links the two sets of scores.

Students were also asked to provide written justification for the highest and lowest number of points awarded. It was noticeable that even in cases where compelling reasons were given to justify a low score, the actual number of points awarded did not deviate far from the 10 points average. For example, missing group meetings, a lack of focus, and not caring about the work were all cited as reasons for awarding a lower score, but in each case 9 points were awarded. Indeed several students, when pressed for the reason why they gave a score lower than 10, responded by writing that they had no choice given the instructions (which prohibited uniform distributions) , and that otherwise they would have awarded 10 points to each team member. Based on the above analysis, the hypothesis was formed that team members will always show a tendency towards the uniform distribution of discretionary points, rather than distinguishing their allocations based on a measured criterion such as ability. Consequently the experiment described in Section 3.1 was designed in order to replicate the results, while overcoming confounds described in Section 3.2.

### *4.2 Results from the Laboratory*

The behavior of the participants in the three experiments is best described by the percentage of participants dividing the discretionary cash amount ($20) uniformly among their four team members. In the first experiment, where each participant was unaware of the identity of their randomly selected partners, and based their judgment solely on the scores from the Raven test, 56% awarded a uniform distribution of $5 per team member.

Table 5 shows the Tobit model results for the first experiment. It is striking that even when participants know nothing about their team members other than their scores from the Raven test, the discretionary cash is not distributed in a way that is statistically significant when we look at the correlation with the Raven scores (t = 1.33, p = 0.198).

While we could regress on many other measured variables, it would be false to do so (and they are hence omitted from Table 5), given that participants did not know the sex, ethnicity etc. of their team members. Not surprisingly though, they prove to be statistically insignificant.

**Table 5 – Tobit Model Results from Experiment 1**

| Variable | Coefficient | Standard Error | t | P>\|t\| | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| (Constant) | 1.36 | 2.74 | 0.50 | 0.625 | -4.30 | 7.03 |
| Quiz | 0.13 | 0.98 | 1.33 | 0.198 | -0.07 | 0.33 |

In the second experiment, where teams were formed from participants who were initially strangers, the percentage awarding the discretionary amount uniformly rose to 88%. The fact that so few differentiated the amounts given to their group members  indicates that none of the now observable characteristics such as gender and race played a significant role in the rationale of the participants, and that intelligence (as measured by the Raven test) was similarly discounted. This is confirmed by the Tobit model results shown in Table 6. In particular the scores from the initial Raven test were not a significant predictor of how the discretionary cash was distributed (t = -0.33, p = 0.743). Indeed the negative t value indicates a slight (though statistically insignificant) propensity among participants to award less to those who scored more on the Raven test.[16]

**Table 6 – Tobit Model Results from Experiment 2**

| Variable[17] | Coefficient | Standard Error | t | P>\|t\| | 95% Confidence Interval | | Tolerance[18] | VIF |
|---|---|---|---|---|---|---|---|---|
| (Constant) | 6.33 | 2.46 | 2.57 | 0.018 | 1.20 | 11.47 | | |
| Female | -0.16 | 0.26 | -0.62 | 0.539 | -0.70 | 0.38 | 0.62 | 1.61 |
| Non-White | -0.11 | 0.28 | -0.39 | 0.70 | -0.71 | 0.48 | 0.51 | 1.94 |
| Risk | -0.09 | 0.08 | -1.18 | 0.253 | -0.26 | 0.07 | 0.87 | 1.14 |
| Raven score | -0.03 | 0.09 | -0.33 | 0.743 | -0.23 | 0.17 | 0.71 | 1.41 |
| Wealth | 0.00 | 0.03 | 0.05 | 0.959 | -0.06 | 0.06 | 0.59 | 1.69 |

The demographic variables included in Table 6 are typical in such experiments. With regard to the distinction between the sexes, Camerer [2003] states that "It is widely thought that women are more likely to sacrifice their own interests for the sake of preserving harmony in relationships, whereas men are more competitive and apply moral principles that override personal relations."

---

[16] Had all of the participants distributed the $20 uniformly among their group members, then all the B weights in the regression analysis would equal zero, so while it appears from Table 6 that there is a slightly negative correlation between the cash distributed and the scores from the Raven test (with each correct answer reducing a participant's amount by 8.3 cents), this apparent anomaly is caused by the small percentage (12%) who did not distribute uniformly, and is not statistically significant (p = 0.421).

[17] The Female and Non-White variables are binary dummy variables. Risk is measured by the number of occasions (with a maximum of 10) that participants chose the risky option when playing the Holt-Laury game. Raven Score indicates the number of questions answered correctly (with a maximum of 30). Wealth denotes the amount of money gained by participants prior to being asked to distribute $20 among their group members.

[18] The collinearity statistics are shown to illustrate that no two exogenous variables are correlated. A VIF figure above 10 (or tolerance below 0.1) would cause concern.

Race was included given that past research has indicated that it sometimes has a stronger effect than gender (for example Eckel & Grossman [2001], who studied ultimatum games). Intelligence is considered to be measured by the Raven Test. Age was not considered given that all the participants were undergraduate students. Attractiveness was not measured, as although Phillips et al. [2008] concluded that altruistic acts make men more attractive to women, there is little evidence that attractiveness leads to altruism.

The $R^2$ value was calculated to be 0.0475, which indicates that even when considering all of the exogenous variables, they only explain 4.75% of the variation observed in the discretionary amounts given to participants. When we just regress the cash amounts based on intelligence (as simulated by the Raven test), $R^2$ drops to 0.003.

For the third and final experiment, the teams were formed by the participants in advance, based on familiarity. Again, the discretionary amounts were awarded with no regard for observable characteristics (see Table 7), with 88% distributing the $20 amount uniformly among their four team members[19].

[19] In addition to familiarity, it is natural to suggest that another latent variable that may account for some of the increase in the percentage of participants awarding a uniform distribution in Experiments 2 & 3 is "effort", given that all the teams were observed by the experimenter working diligently and cooperatively on the second set of questions from the Raven test. This could also explain the increased percentage when compared with Experiment 1, where team members did not get to meet each other. However work done in this area (see Rutström & Williams [2000]) suggests that the effect of effort on distributions is not as significant as that of wealth, which was not found to be a significant predictor here (see Tables 6 & 7).

**Table 7 – Tobit Model Results from Experiment 3**

| Variable | Coefficient | Standard Error | t | P>\|t\| | 95% Confidence Interval | | Tolerance | VIF |
|---|---|---|---|---|---|---|---|---|
| (Constant) | 5.18 | 1.38 | 3.77 | 0.001 | 2.31 | 8.05 | | |
| Female | 0.21 | 0.26 | 0.84 | 0.411 | -0.32 | 0.75 | 0.66 | 1.51 |
| Non-White | -0.35 | 0.27 | -1.30 | 0.207 | -0.91 | 0.21 | 0.68 | 1.48 |
| Risk | -0.12 | 0.93 | -1.29 | 0.211 | -0.31 | 0.07 | 0.96 | 1.04 |
| Raven score | 0.04 | 0.05 | 0.80 | 0.431 | -0.07 | 0.15 | 0.64 | 1.57 |
| Wealth | -0.02 | 0.03 | -0.58 | 0.565 | -0.08 | 0.05 | 0.56 | 1.79 |

The $R^2$ value of 0.109 in Experiment 3 was again extremely low, and indicative of the fact that very little of the manner in which the discretionary cash was distributed can be explained by the measured variables.

## 5. CONCLUSION

There are some who will never be convinced that pure altruism exists. Sober and Wilson [1998] jokingly contend that "Even saints could be regarded as selfish if they perceive their lives of sacrifice as tickets to heaven."

However the results of the laboratory experiment confirm that none of the standard demographic or latent control variables explain the overwhelming propensity of participants to uniformly distribute discretionary money. Given the extremely low $R^2$ values, it is doubtful that the results could be statistically explained using an alternative, more complicated, model. This is in spite of the data being extremely suggestive, and hence it is reasonable to assume that by spending time together in groups, and decreasing the social distance between them, the members of a team form a bond similar to that seen in families, and that therefore the theories of kin altruism, a subject considered at length by Hamilton and others in the 20th century (see Section 1.2), apply in this setting[20].

While it is right to question whether the apparent altruism shown by participants extends to more diverse populations, and that the external validity of the findings suffers to the degree that college students do not mirror them, the strength of the laboratory data makes it likely that the conclusions reached can be generalized, and is the natural direction of any further study.

[20] It is natural to interpret the uniform distribution of discretionary points/cash as egalitarian rather than altruistic. However such a tendency should lead to a negative correlation between intelligence/wealth and the amount of points/cash received, in order to correct the inequality between participants created by earlier parts of the experiment. A more appropriate explanation is that the distributions are in line with the "maximin" concept proposed by Rawls [1973], whereby participants derive utility from others getting value and by ensuring that the lowest amount given is maximized. This has been incorporated into models of altruism considered by Cox [2007].

## REFERENCES


Andreoni, J. [1990]. Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. *Economic Journal, 100: 466-477.*

Andreoni, J. & Miller, J. [2002]. Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism. *Econometrica*, *70(2): 737-753.*

Axelrod, R. [1984]. *The Evolution of Cooperation*. New York, NY: Basic Books.

Axelrod, R. [1986]. An Evolutionary Approach to Norms. *American Political Science Review*, *80: 1095-1111.*

Batson, C.D., & Shaw, L.L. [1991]. Evidence for Altruism: Toward a Pluralism of Prosocial Motives. *Journal of Psychological Inquiry, 2(2): 107-122.*

Becker, G. [1974]. A Theory of Social Interactions. *Journal of Political Economy, 82: 1063-1093.*

Binmore, K. [1998]. *Just Playing: Game Theory and the Social Contract II*. Cambridge, MA: MIT Press.

Binmore, K. [2006]. The Origins of Fair Play. *Papers on Economics & Evolution, 0614*. Jena, Germany: Max-Planck-Institut für Ökonomik.

Buchanan, J.M. [1975]. The Samaritan's Dilemma, in E. Phelps (ed.), *Altruism, Morality and Economic Theory*. New York, NY: Russell Sage Foundation.

Camerer, C.F. [2003]. *Behavioral Game Theory*. Princeton, NJ: Princeton University Press.

Cherry, T., Frykblom, P., & Shogren, J. [2002]. Hardnose the Dictator. *The American Economic Review*, *92 (4): 1218-1221.*

Coate, S. [1995]. Altruism, the Samaritan's Dilemma, and Government Transfer Policy. *American Economic Review*, *85(1): 46-57.*

Comte, A. [1851, 1969-70]. *Système de politique positive: Oeuvres d'Auguste Comte Tomes 7-10.* Paris: Editions Anthropos.

Connolly, K. & Martlew, M. [1999]. *Psychologically Speaking: A Book of Quotations.* Leicester, UK: BPS Books.

Cox, J.C. [2007]. Trust, Fear, Reciprocity, and Altruism: Theory and Experiment, in Sobei H. Oda (ed.), *Developments on Experimental Economics: New Approaches to Solving Real-world Problems.* Berlin, Germany: Springer-Verlag.

Croson, R.T.A. [1999], Contributions to Public Goods: Altruism or Reciprocity? Working Paper #96-08-01, University of Pennsylvania.

Dawkins, R. [1976]. *The Selfish Gene*. Oxford, UK: Oxford University Press.

Eckel, C. & Grossman, P. [2001]. Chivalry and Solidarity in Ultimatum Games. *Economic Inquiry, 39(2): 171-188.*

Fehr, E. & Gächter, S. [2000a]. Fairness and Retaliation: The Economics of Reciprocity. *Journal of Economic Perspectives, 14 (3): 159-181.*

Fehr, E. & Gächter, S. [2000b], Cooperation and Punishment in Public Goods Experiments. *American Economic Review*, *90: 980-994.*

Forsythe, R., Horowitz, J., Savin, N.S., & Sefton, M. [1994]. Fairness in Simple Bargaining Games. *Games and Economic Behavior, 6: 347-369.*

Gintis, H. [2000]. Strong Reciprocity and Human Sociality. *Journal of Theoretical Biology, 206: 169-179.*

Grether, D.M. & Plott, C.R. [1979]. Economic Theory of Choice and the Preference Reversal Phenomenon. *American Economic Review, 69 (4): 623–638.*

Güth, W., Schmittberger, R. & Schwarze, B. [1982]. An Experimental Analysis of Ultimatum Bargaining. *Journal of Economic Behavior and Organization, 3: 367-388.*

Hamilton, W.D. [1964]. The Genetic Evolution of Social Behavior I-II, *Journal of Theoretical Biology*, *7: 1-52.*

Harrison, G.W. [2006]. Experimental Evidence on Alternative Environmental Valuation Methods. *Environmental and Research Economics, 34: 125-162.*

Harrison, G.W. & Johnson, L.T. [2006]. Identifying Altruism in the Laboratory, in D. Davis and R. Mark Isaac (ed.), *Experiments in Fundraising and Charitable Contributions.* Greenwich, CT: JAI Press, Research in Experimental Economics, 11: 177-223.

Hoffman, E., McCabe, K., & Smith, V. [1996]. Social Distance and Other-Regarding Behavior in Dictator Games. *American Economic Review, 86 (3): 653-660.*

Holt, C.A. & Laury, S.K. [2001]. Risk Aversion and Incentive Effects. *American Economic Review, 92 (5): 1644-1655.*

Kahneman, D., Knetsch, J.L., & Thaler, R. [1986]. Fairness as a Constraint on Profit Seeking: Entitlements in the Market. *American Economic Review, 76 (4): 728-741.*

Maynard Smith J. [1988]. *Games, Sex, and Evolution*. New York, NY: Harvester-Wheatsheaf.

Phillips, T., Barnard, C., Ferguson, E., & Reader, T. [2008]. Do Humans Prefer Altruistic Mates? Testing a Link between Sexual Selection and Altruism towards Non-Relatives. *British Journal of Psychology, 99: 555-572.*

Rapoport, A. [1991]. Ideological Commitments and Evolutionary Theory. *Journal of Social Issues, 47: 83-99.*

Rawls, J. [1973]. Distributive Justice, in E.S. Phelps (ed.), *Economic Justice: Selected Readings*. Harmondsworth, UK: Penguin Education.

Rutström, E.E. & Williams, M.B. [2000]. Entitlements and Fairness: An Experimental Study of Distributive Preferences. *Journal of Economic Behavior & Organization,43: 75-89.*

Salsman, R. [2009]. Altruism: The Moral Root of the Financial Crisis. *The Objective Standard: A Journal of Culture and Politics, 4(1).*

Scott, N. & Seglow, J. [2007]. *Altruism*. Maidenhead, UK: Open University Press.

Sober, E. & Wilson, D.S. [1998]. *Unto Others*. Cambridge, MA: Harvard University Press.

Solow, R.M. [1971]. Blood and Thunder. *The Yale Law Journal, 80 (8): 1696-1711.*

Stark, O. [1989]. Altruism and the Quality of Life. *American Economic Review, 79: 86-90.*

Stark, O. [1995]. *Altruism and Beyond.* Cambridge, UK: Cambridge University Press.

Stone, D. [2008]. *The Samaritan's Dilemma*. New York, NY: Nation Books.

Sugden, R. [1982]. On the Economics of Philanthropy. *Economic Journal, 92: 341-350.*

Sugden, R. [1984]. Reciprocity: The Supply of Public Goods through Voluntary Contributions. *Economic Journal, 94: 772-787.*

Titmuss, R. [1970]. *The Gift Relationship: From Human Blood to Social Policy*. New York, NY: Pelican Books.

Trivers, R.L. [1971]. The Evolution of Reciprocal Altruism. *Quarterly Review of Biology, 46: 35-57.*

Upton, W. [1974]. Altruism, Attribution and Intrinsic Motivation in the Recruitment of Blood Donors, in *Selected Readings in Donor Motivation and Recruitment.* Washington, DC: American Red Cross.