The Pennsylvania State University

The Graduate School

College of Information Science and Technology

MORAL DECISION MAKING UNDER MONETARY CONSIDERATIONS

A Thesis in

Information Science and Technology

by

Chenrong Qin

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

May 2019

The thesis of Chenrong Qin was reviewed and approved* by the following:

David Reitter
Associate Professor of Information Sciences & Technology
Thesis Advisor

Tinghao Huang
Assistant Professor of Information Sciences & Technology

Sarah Rajtmajer
Assistant Professor of Information Sciences & Technology

Mary Beth Rosson
Professor of Information Sciences & Technology
Associate Dean, College of Information Sciences & Technology
Head of Graduate Program

*Signatures are on file in the Graduate School.

Abstract

A moral decision does not always mean rational decision. The moral decision is a choice made based on a person's feeling and what they believe is proper behavior. We use monetary choices to explore how group size is correlated with the amount of benefit gained by the subject compared to the harm that their decision inflicts on others. To test this behavior, we measure how Amazon Mechanical Turk workers make choices in a game similar to the Dictator Game where their decisions affect simulated participants.

Preliminary work shows that people are more likely to accept inflicting costs to the group if they are offered more money or if the group loses less money per person. My thesis explores how people evaluate the morality of gaining money at a cost to others depending on the amount of personal gain, the cost to others, and the number of people affected. Our experiments will also reveal how individuals behave when making decisions about different sized groups.

We conduct a series of experiments using Amazon Mechanical Turk and Qualtrics while each experiment kept one of the independent variables constant. To the data from each experiment, I fit classification model, multiple logistic regression model and generalized linear mixed model.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to give special thanks to my advisor, Associate Professor David Reitter, for his encouragements and useful advice. He is both my mentor and good friend. He has been exceptionally cooperative from the day I joined her research lab. He is always willing to discuss with me and help me explore my interesting field. Next, I would like to thank my co-advisor, Matthew Kelly, for providing valuable advice with the experiment design.

I would also like to thank my friends from the College of Information Science and Technology for their assistance and encouragement.

Chapter 1

Introduction[1]

A rational model is usually with respect to an entire goal. Based on models of decision making, a decision maker trying to maximize personal benefit should ignore the fact that they would be harming others at all times for the sake of their own personal gain. My research seeks to answer the following questions: If an individual is presented with the opportunity to give or take certain amounts of money from a group, will the individual engage in the same thought process as if that individual were not an individual, but a group? Will the individual make a different decision if they are affecting only one other individual? Is it possible that the size of the group results in a different decision made on the part of the individual?

It is important to note that the test subjects do not know the identity of the individuals in the group affected by their decisions. We know that people who know each other are less likely to take money from others. In our experiments, we want to investigate how people make decisions when people do not know each other. This is acceptable since previous research has shown that, when presented with the choice to harm or benefit different groups whose members were unknown to the test subjects, it was found that the choices were random (Baron, 1995). This previous research was not concerned with different sizes of groups; therefore, our study will reveal how group size is correlated with the amount of harm or benefit, which is monetary in this case.

This research explores how people make decisions when facing monetary considerations that pit the interests of the group against benefit for oneself. For example, given a choice to either earn $100 at a cost of $1 to five other people, or to earn nothing but cost those five people nothing, most people would

[1] Some of chapter 1 is from a term paper for IST 597 Special Topics: Decision making, which was co-authored with David Reitter, Chaoran Chen, Jessika Kinley Turner, Jiadi Liu, Nasim Motalebi, Jacob Oury, Sayali Phadke, Yasmin Tantawi.

probably choose to earn $100. On the other hand, given an offer to earn $1, while the other five people lose $100, most people would probably refuse the offer. Our major interest is in determining the impact of individual versus group loss and correlation of personal gain and the victim's losses. Does a decision-maker prefer to inflict monetary losses on others for a monetary gain that is less, equal, or more than the loss of the victims? Do decision makers prefer to inflict monetary harm on individuals or prefer to harm a group?

The experiment, which will be described in more detail in subsequent sections of this thesis, examines the tradeoffs involved in profiting from other people's losses. Our experiment will ask a number of participants whether they were willing to tax a group of people that would lose a variable amount of money for one's own gain. The experiment we describe here examines the tradeoffs involved in profiting from other people's losses. We will model the acceptability of such tradeoffs in face of changing group size. Is it more acceptable to tax many people a little, or to tax few individuals a lot, in order to derive a personal gain? This experiment will quantify the amount of possible money for which individuals are willing to take or give money to others and reveal how individuals behave when their decision affect different-sized groups.

Chapter 2

Literature Review[2]

Individuals are more prone to experience negative feelings as a result of a loss than positive feelings as a result of a gain, however, their feelings are not directly proportional to the amount of gains or losses. Rather, it has been found through numerous experiments that the greater the amounts of objects one has, be they monetary wealth, objects of material value, or objects of sentimental value, the less happiness one perceives as the amount of these objects increases (Kahneman, 1979). It has also been revealed that individuals act more rationally when they work in groups rather than individually (Kugler et al., 2012). In addition, it has also been found that individuals are more likely to contemplate severely harming one or more individual if that harm, including fatal harm, is produced via an impersonal switch; whereas these same individuals are much less likely to harm one individual if that harm would be produced directly by their own hand (Shallow et al., 2011).

Brief et al. (2001) explain moral decision making as an act of engaging with a decision that requires a choice between two or more important values. They bring up an example in which a person holds dear both the values of equality and a comfortable life and is confronted with the question ‘Should taxes be raised to aid the poor?’. It can then be said that the person’s answer to this question is a product of moral reasoning. The answer to this question relies on a complex of personal and contextual factors which would affect the moral decision-making process.

In our study, we are interested in understanding moral decision-making in a condition in which a person could inflict monetary harm on others for personal monetary gain. Our question pertains the issue of

---

[2] Some of chapter 2 is from a term paper for IST 597 Special Topics: Decision making, which was co-authored with David Reitter, Chaoran Chen, Jessika Kinley Turner, Jiadi Liu, Nasim Motalebi, Jacob Oury, Sayali Phadke, Yasmin Tantawi.

whether personal gain justifies harming others and if it does, when or what triggers it. Our major interest is the impact of individual versus group loss, and the relationship between personal gain and the victim's losses. This relationship could take many forms. For one, we are interested in understanding whether a decision maker would inflict monetary harm on others for a monetary gain that is less, equal, or more than the monetary loss of the victims. Secondly, would a decision maker be reluctant to harm an individual and prefer to harm a group instead? Or vice versa: would a decision maker prefer to harm an individual rather than a group? Finally, we would like to understand if putting the decision maker at the risk of being harmed by their victims would impact their moral decision-making behavior. In the existing literature on moral decision making some of such factors have been studied or identified separately. But the literature has failed to study the impact of the complex set of factors on moral behavior. We will add these factors into our experiments.

Amir et al. (2016) find that people cheat on groups more than they cheat on individuals as it feels that less harm would be caused to individuals in a group rather than an individual alone. However, Amir et al. do not address the impact of the amount of personal gain in the moral decision-making process. Therefore, in this study we would like to understand whether the amount of personal gain would impact a moral decision in harming a group versus harming an individual. Is there a flipping point that moral decisions would be ignored for the sake of personal gains regardless of the amount of harm that is inflicted upon others? In other words, what amount of personal gain would make an individual inflict harm onto others regardless of the severity of the harm?

A similar question is studied by Trevino & Youngblood (1990), where the authors have conducted experiments on the impact of rewards in unethical decision-making. Their results indicate that extrinsic rewards (such as monetary rewards) significantly increase unethical decision-making. We would like to expand their studies in order to understand the correlation between the amount of reward that people are promised for an immoral behavior and the amount of harm that they would cause to others by making that decision. More specifically, would an individual would be willing to cause monetary harm to another person or a group if their personal monetary gain (reward) is equal, more, or less than the monetary loss of others? In addition, we would like to know how unethical behavior changes if the decision maker is also exposed to potential harm from those who s/he is harming? In sum, we are motivated to study the impact of the following complex set of factors on immoral decision making: the number of people that would be harmed (individual versus a group), the amount of harm to others in relation to the amount of personal rewards, and the amount of potential harm that would be caused by the decision.

We argue that the identity of the victims and the decision makers must remain anonymous since it would better lead to immoral behavior. Based on a study by Gino et al. (2010), a decision is considered to be unethical when the victims are identifiable. Therefore, we aimed to keep the identity of the victims and the decision maker unidentifiable to encourage unethical behavior.

Based on previous research, we designed a set of experiments in which all such factors could be studied in a single, simple game. While the complexities of moral decision making can be studied in the brain using neuroimaging and FMRI (Greene et al., 2004; Heekeren et al., 2003), we propose a strictly behavioral study with the aim of improving the theoretical understanding of the psychology of moral decision-making behavior. The next section is dedicated to describing the design of the game, followed by the experiment itself, the results, and the behavioral model which was concluded from the experiment results.

Chapter 3

Research methodology

Chapter Overview

There are three main sections in this chapter. In the first section, I summarize how I designed a survey study using Qualtrics and describe how I recruited participants in Amazon Mechnical Turk. In the second section, I elaborate on the data collection process, in particular describing on how I conducted the survey and gathered meaningful information. In the third section, I describe the details of the survey questions.

Experimental Design

To conduct our experiment on moral decision-making, we want to put people in a position where they would need to cause monetary harm to others in order to receive their monetary rewards. As we mentioned before, a rational model depends on a goal. In our case, if my goal is to finish the survey quickly or make the most profit, I will always click yes all the time. If my goal is to benefit the other group, I will consider that whether my personal gain is more than group loss. If my goal is to give away money to others, I would choose no in all questions. Our goal was to stimulate a real gamble and study the effect of various factors that would encourage moral behavior (preventing harm to others) and factors that would cause immoral behavior (personal gain at a cost to others) on decision making process. We put the decision maker in a position of potential loss as well. In other words, their decisions would harm others, and other's decisions would harm them. We hypothesized that people would not always choose to take the money because of a sense of empathy towards those who they are inflicting a monetary loss upon. We also hypothesized that decision makers would prefer to inflict losses on individuals as opposed to a group of people.

Next, we hypothesized that the amount of monetary gain and its relationship to the monetary loss of others would impact the decision maker's moral behavior. As a result, I set three main variables in which Y is the total monetary gain of the decision maker, N the number of victims in a group, and X the amount of loss per person. The experiments place the subjects into a situation where they are forced to decide if they want to increase their own gain at the expense of other simulated participants. The subject is aware that their gains come directly at the loss of one or several other participants. Each scenario is based around some number of the other participants being forced to give up some portion of their earnings if the subject chooses to accept the forced gamble.

We conducted three experiments with different ranges of values. In the first experiment, we keep N*X fixed to 5 (i.e., the product of the individual loss by group size is constant). In the second experiment, X (the individual loss) is fixed to 1. In the third experiment, Y (the personal gain) is fixed to 1. Based on

first two experiments, I find that N (group size) is not significant from range 5 to 10, so I decrease the range of N to {1,2,5} in the third experiment. Experimental design is shown below in Table 1.

1st experiment (93 participants)	2nd experiment (86 participants)	3rd experiment (97 participants)
N*X is fixed to 5 N = {1,5,10} X = {0.5,1,5} Y = {1,5,10}	X is fixed to 1 N = {1,5,10} X = {1,1,1} Y = {1,5,10}	Y is fixed to 1 N = {1,2,5} X = {0.5, 1,2} Y = {1,1,1}

Table 1: The experimental conditions used for the game in three sub-experiments.

The original design of three experiments are not good enough since we only varied the values of each predictor but the overall design looks unreal. People cannot consider questions in a moral way. For the fourth experiment, we made changes to the experimental design in order to encourage participants to consider the effect of group size and the morality of their decisions, as detailed in the *Procedure*. The new experimental conditions for the fourth experiment are shown in table 13.

4th experiment (95 participants)	N = {1,5,10} X = {0.5,1,2} Y is fixed to 5

Table 13: Experimental conditions used for the game in the 4th experiment.

Participants and data collection

The participants were recruited from Amazon Mechanical Turk and an online questionnaire was presented to them using Penn State Qualtrics. After I got two datasets from both MTurk and Qualtrics, I combined these datasets by connecting each one's unique MTurk code to get the data for each participant.

The data collected from each experiment will consist of the demographic information and the subject responses to each question. Standard descriptive statistics will be used to characterize trends and relationships between the demographic data and responses. We will explore possible connections between number of other participants, the relative value of the subject's bonus compared to the loss by other subjects, and decisions made by each subject.

A total of 276 subjects (Male = 192, Female = 84) were tested in the first three experiments. Participants ranged in age from 19 to 65 (Mean = 32). Most participants were from the United States (n = 164). A large portion were also from India (n = 90). The other countries represented are Andorra, Belize, Australia, Germany, and Philippines.

In order to improve the quality of survey, we added two more restrictions in the fourth experiment. HIT approval rate for all requesters' HITs are more than 95% and location is the United States. Workers should be in the United States so we can avoid the problem of income difference. A total of 95 subjects (Male = 67, Female = 28) were tested in the fourth experiment. Participants ranged in age from 19 to 67 (Mean = 34). The total number of decision 'yes' is 691 and 'no' is 155. Four people answered all 'no's and 42 people answered all 'yes'. Below is the table of frequencies for each condition of experiment, or 'case'.

Decision	case1	case2	case3	case4	case5	case6	case7	case8	case9	Total
NO	11	8	12	6	12	23	19	31	33	155
YES	83	86	82	88	82	71	75	63	61	691

Table 14: Number of decisions for each case

Each individual gets a bonus when they finish, which improves their motivation to do the questionnaire. For Experiments 1, 2, and 3, the bonus is 2 cents. The results were not good enough, so I changed the bonus to a raffle ticket in Experiment 4. Data from participants who finish too quickly (i.e., in less than two minutes) or people who do not pass the attention check question are not included in the analysis and we would not approve for them. Compared to doing surveys on a campus where most participants are students and the age range is very narrow, MTurkers have a larger range of ages that is more representative of the population.

Procedure

In the procedure part, I want to separate two experiment designs because we almost change all the small details in the fourth experiment to improve the quality of survey. More details will be compared between these two experiments.

Each subject begins the experiment by accepting the experiment via Amazon Mechanical Turk. Next, the subject follows a link to Qualtrics where the experiment is hosted. The subject then fills out their demographic information (i.e. age, gender, country of origin). The instructions for the first three experiments are below:

> You are involved in a gambling game with 50 other random Amazon Turkers. In each round, you will be presented with a piece of information. You will then use this information to decide whether or not to participate in the game.
>
> You will gain 0.01 dollars for completing each round and you can gain additional bonus depending on your answers. However, your decision in each round could potentially add to your bonus by taking away the bonus from other Turkers. Keep in mind, others will be playing a forced gamble on you too. The other players will also be presented with information that could also affect the gamble and your gain.
>
> Your own reward will be revealed to you at the end of this game. However, your final bonus will be tallied after all the other Turkers compete in this game and will be applied within 36 hours of the HIT concluding.

After reading these instructions, the subject begins the experiment. For each experiment, the subject is presented each of the nine scenarios in random order and asked to respond either “yes” or “no” to each scenario before continuing to the next. An example prompt for a scenario is shown below. This is the prompt for the [N = 1, X = 2, Y = 1] condition. The participant will gain ¢1 while the other participant will lose ¢5. After the subject responds to the nine scenarios, they are presented with the total bonus earned and informed that their own bonus will be used for other subjects’ responses. The maximum bonus able to be earned is ¢30.

> Read this gamble carefully. This is a real gamble and other Turkers will also be gambling against you. Your responses in each round will affect other Turkers’ bonuses, and their responses will affect your bonus.
>
> One participant has bet 5 cents in this round. You know that if you accept their bet, they will lose their 5 cents and you will gain 1 cent. Are you willing to accept their bet for a 1 cent gain? You are guaranteed a 2 cents' bonus for answering this question, regardless of your choice.

Afterwards in the fourth experiment, we tried to improve the experiment design because the previous results showed that no matter in which condition, there are a large proportion of people who always answered yes. I believe the previous design has some problems. The two basic reasons are: 1. Mturk workers can communicate within the forum so they may know each other and realize this is not a real game. 2. The bonus for each question is only 2 cents. That means participants will not have much motivation to complete the survey morally. They mostly consider their monetary gain even it is very small amounts, so that they cannot connect the game to moral considerations. The new instructions are below.

> You will gain 50 cents for completing the survey and you can participate in a raffle to win an **additional bonus**. Each Turker starts the game with a raffle ticket for a prize with an initial value of **$2**. Your raffle ticket has a 15**%** chance of winning. The value of your raffle prize can be **increased** by your decisions and **decreased** by other Turkers' decisions.
>
> There are nine rounds in this game. In each round, you will have the choice to decrease the value of other Turkers' raffle prizes to increase the value of your own raffle prize. But other Turkers' will be given the same choice about your raffle prize.

The goal of our fourth experiment is to improve the quality of survey, so we changed most of settings and increase some small details to simulate the game more real. The biggest change is that we changed the monetary gain in each scenario to an increase in the value of the raffle ticket by $0.50. This step is very delicate because we construct an environment which people can win a much larger bonus. The reality is that our budget (total bonus) does not increase a lot because we limit the probability of winning to only 15%. We also assigned them an initial ticket value of $2 since we do not want people who say ‘no’ in the first couple of questions to get a negative value. If one participant answers ‘yes’ in one question, the value would add up to $2.5. The maximum raffle prize is $4.35. Compared with previous given maximum bonus of 10 cents, the relative large amount of bonus is more attractive and people will be more willing to take part in the game. Before they start real questions, they are still asked for demographic information like age, gender and country of birth. An example prompt is shown below.

Read the following carefully. Your responses in each round affect other Turkers' bonuses, and their responses affect your bonus.

The value of your raffle prize is currently $.

You are in a group of ten Turkers online right now. The Turkers' usernames are: tommysion, AholicJerry, ChrisChilli, Jameshu, KeepupAholic, BabyJamie, Jimmyqin, SnowBlue, StudioSparkling, smugaboo.
You can decrease the raffle prize for 5 other MTurker by $0.10 each. If you do, your current raffle prize will increase by $0.50. Will you accept this offer?

Apart from the raffle ticket function, we can see other changes from the prompt. Questions look more concise and we include some new information. There is also a list of usernames and their decisions provided after each question. We also want to inform workers exactly who take the money from them. Since we asked for usernames at the very beginning of survey, this is a very useful tip related to the question because others' usernames are shown to other MTurkers. They do not need to pick their real usernames, so they can choose any name they want. We randomly picked ten usernames from the list and post into the question. This step is very necessary because we construct a circumstance that they are competing with others.

On the other hand, we add a live wallet to show each one's money gain or loss after each question. It is convenient for people to see the value of raffle ticket live and see how it has been affected by their own answers. For example, if I choose 'no' for first three answers, then my wallet might be empty and I will pick more 'yes' instead. After they complete the survey, they will see their net gain for all questions.

One more important detail is that we add one figure in each question to help people more easily understand the offer. One example figure looks like:

We use circles to represent monetary gains and losses. A half circle represents 5cents and a full circle represents 10 cents. The participant is the white figure and the circles indicate your earnings. The ten blue figures indicate other Turkers and the circles indicate their losses. For example, in this diagram, you are being offered 50 cents at the cost of 5 cents to five other Turkers.

One attention check question is also included within the real questions. The answer should be “No” to get the task approved. Incorrectly answering this attention check led to the excluding of the participant’s data from our study.

After we collected the data, we randomly chose 15 people from the whole datasets and gave them their corresponding raffle ticket earnings and posted the list of winners to a web page that we provided in the survey. Updated results will be shown in the following section.

Chapter 4

Measurement methods

Multiple Logistic regression

The first method we use is multiple logistic regression with a two-way table. In particular, binary logistic regression is a special type of regression where binary response variable is related to a set of explanatory variables. Probability of the response taking a particular value is modeled based on combination of values taken by all predictors.

LDA (Linear Discriminant Analysis) & QDA (Quadratic Discriminant Analysis)

LDA and QDA algorithms are based on Bayes theorem and are different from the Logistic Regression. In Logistic regression, it is possible to directly get the probability of an observation for a class for a particular observation.

LDA (Linear Discriminant Analysis) is used when a linear boundary is required between classifiers and QDA (Quadratic Discriminant Analysis) is used to find a non-linear boundary between classifiers. If the response classes are separable or distribution of observations for all classes is normal, both LDA and QDA methods will work better. LDA and QDA work well when class separation and normality assumption holds true. The more classes are separable and the more the distribution is normal, the better will be the classification result for LDA and QDA (Shekar, 2018).

Naïve Bayes Classifier

Naive Bayes is a simple, yet effective and commonly-used, machine learning classifier. It is a probabilistic classifier that makes classifications using the Maximum A Posteriori decision rule in a Bayesian setting. It can also be represented using a very simple Bayesian network. Naive Bayes classifiers have been especially popular for text classification, and are a traditional solution for problems such as spam detection (Soni, 2018).

Generalized Linear Mixed Model

Another method we chose is the generalized mixed logit model (GLMM), which is also designed for binomially distributed outcomes. The GLMM describes an outcome as the linear combination of fixed effects and conditional random effects associated with subject. In our experiments, fixed effects are N, X and Y, and the random effect is subject. ‘Subject' as a grouping factor is used for within-subject predictors.

Chapter 5

Results and findings

Chapter Overview

In this section, we discuss the results of descriptive analysis, including visualizations and exploratory models, followed by results from the statistical model implemented to analyze different experiments. As we mentioned before, we combine the first three experiments as a whole dataset to analyze since we keep one variable constant in each experiment. Based on results of first three experiments, we conduct one more experiment to further examine the question of the effect of group size. We use a lot of graphs to illustrate the role each variable play in affecting the prediction of the proportion of people who answer yes or no.

Descriptive analysis

In the descriptive analysis, we examine the proportion of participants who agreed to impose a cost X on N people for a personal gain of Y. Figures 1, 2 and 3 with line plots are correspond to three experiments. The y axis (proportion of subjects responding 'yes') is related to different settings of scenarios which takes an input with any value from negative to positive infinity and return an output that always takes values between zero and one. This is interpretable as the probability of the decision yes. In each of these, we hold one of the three variables constant at each level and observe the progression of the proportion of 'yes' responses as the values of the other two variables change.

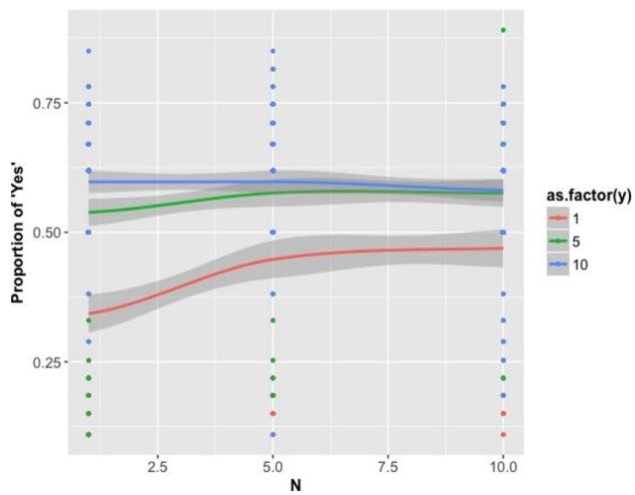

Figure 1-1: Proportion of agreement given number of people affected (N) at each level of personal monetary gain (Y), with SE indicated in grey.

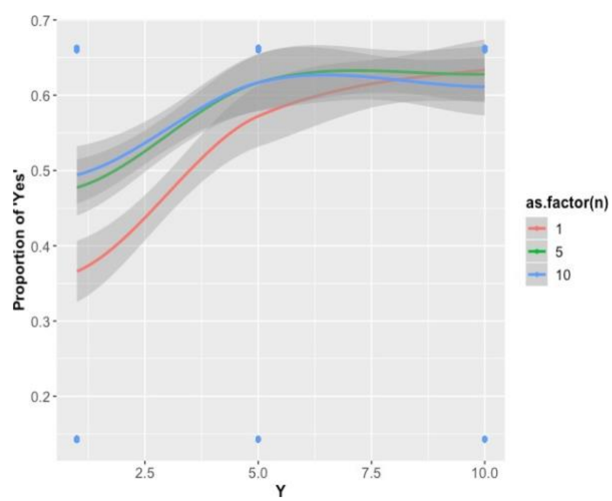

Figure 1-2: Proportion of agreement given personal monetary gain (Y) at each level of number of people affected (N), with SE indicated in grey.

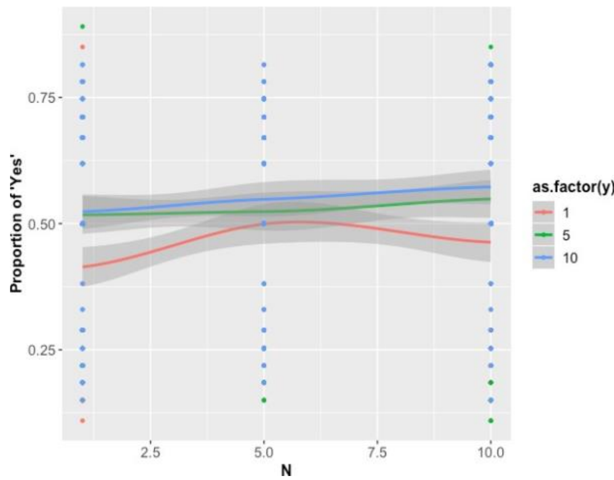

Figure 2-1: Proportion of agreement given number of people affected (N) at each level of personal monetary gain (Y), with SE indicated in grey.

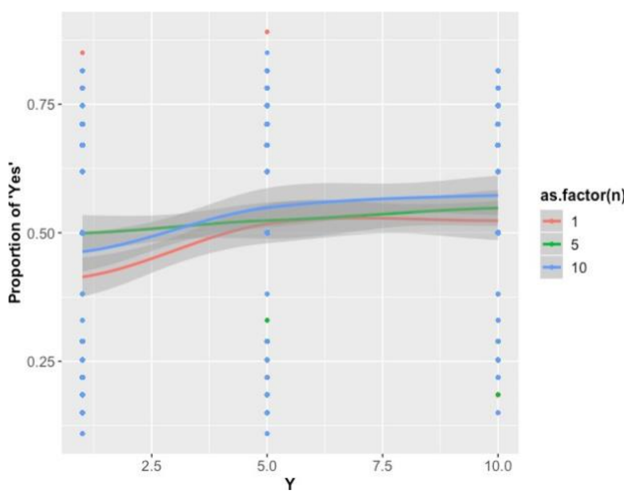

Figure 2-2: Proportion of agreement given personal monetary gain (Y) at each level of number of people affected (N), with SE indicated in grey.

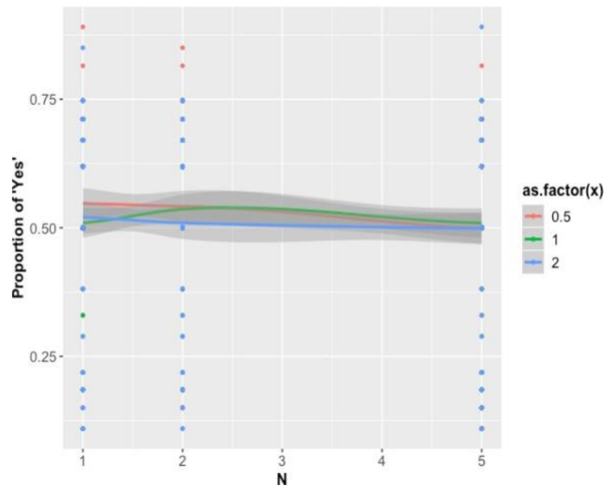

Figure 3-1: Proportion of agreement given number of people affected (N) at each level of cost per person(X), with SE indicated in grey.

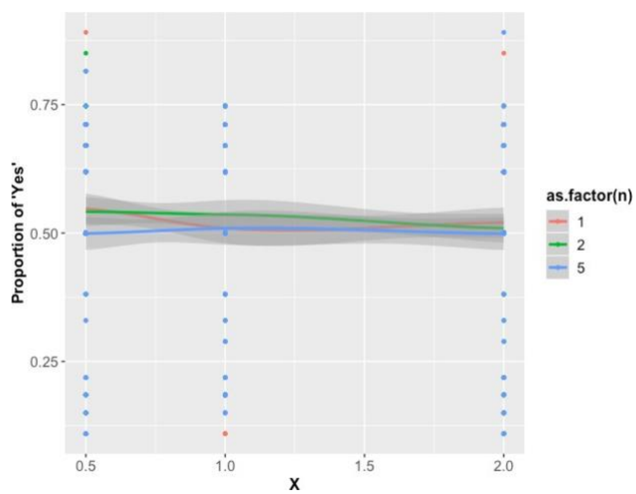

Figure 3-2: Proportion of agreement given Cost per person (X) at each level of number of people affected (N), with SE indicated in grey.

In all experiments, continuous variables such as N and Y were studied only at discrete values. From Figure 1 (part 2), it's clear that the slope, from when Y increases from $1 to $5 to when it increases from $5 to $10, is different at each level of N. Afterwards we found the slope change for N was not as drastic as in the other cases. We notice that where the personal gain (Y) is less than or equal to the loss per person (X) multiplied by the number of other participants (N) ($Y < N * X$ and $Y = N * X$), the proportion of 'yes' responses is higher as Y increases. Whereas, when $Y > N * X$, the proportion of 'yes' responses reduces as Y increases.

In Figures 2 and 3, it is hard to see the slope difference because the three lines are very close. We can only observe that when Y or X are in different levels, the slope of N is totally different in each experiment (see Figures 1-1, 2-1, 3-1), so that we cannot predict whether N has a main effect. On the other hand, we can see that as N is kept constant, the proportion of decision 'yes' increases when Y increase or X decreases. Since these lines are not linear, so there should be some non-linear effects (i.e. quadratic effect) that are generalizable. Furthermore, we would need to conduct a study where the experimental conditions are specified at more values of these variables.

It is important to note here; each experiment has different settings and we try to keep one variable constant in each. In particular, for the experiment 1, the product of N and X is constant, making those two variables perfectly collinear. As a result, in order to fit the best model, next step is to combine all three datasets together.

For the fourth experiment, we add more descriptive analysis graphs to analyze the data. This time we only varied variables X and N since we found that Y is the most important predictor based in the first three experiments. The results are illustrated in Figure 4.

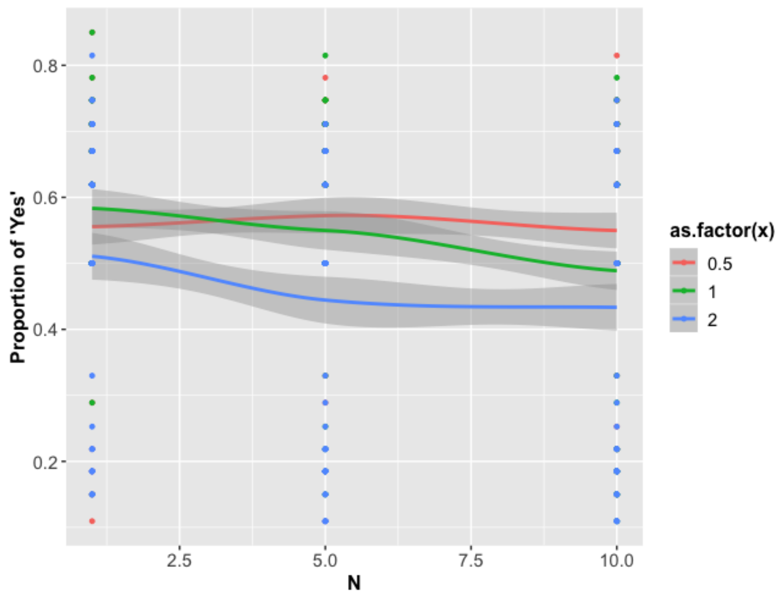

Figure 4-1: Proportion of agreement given number of people affected (N) at each level of cost per person(X), with SE indicated in grey.

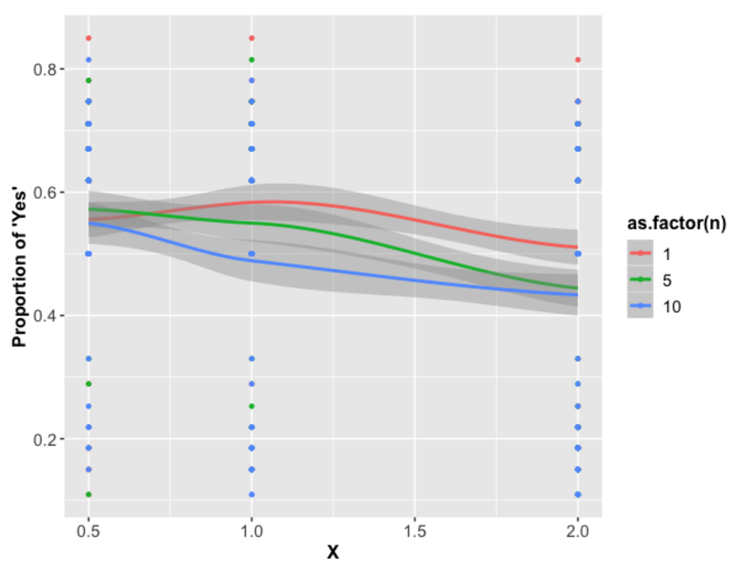

Figure 4-2: Proportion of agreement given Cost per person (X) at each level of number of people affected (N), with SE indicated in grey.

Figure 5 shows the proportion of ‘yes’ responses for each unique combination of N and X and includes bootstrap error bars for a 95% confidence interval. The differences noticed in this visualization are tested using some classification models on the individual Yes-No responses with X and N as continuous explanatory variables, including their main effects and the interaction term.

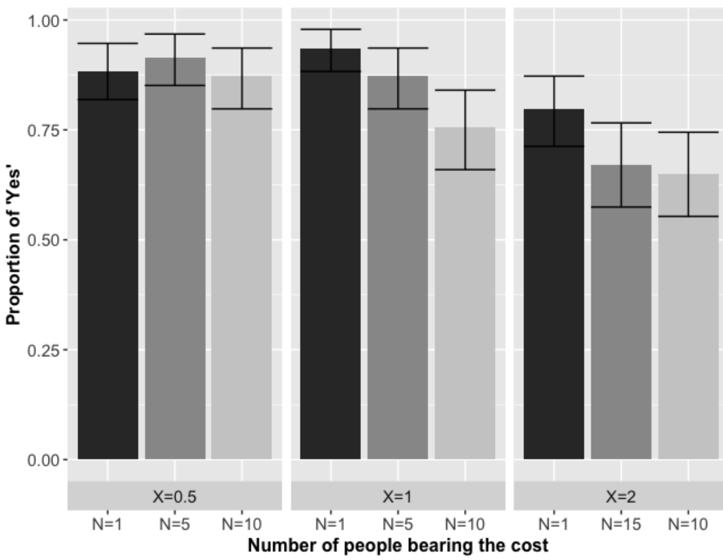

Figure 5: Proportion of agreement for cost per person(X), as grouped by the number of other people(N), with bootstrapped 95% confidence interval

In Figure 5, we notice that, overall, the higher the cost per person(X), the less likely participants are to impose a cost upon others for their own benefit, holding the number of others the same. This is also true for N. Keeping X constant at each level, as the number of other people increases, individuals are less likely to accept the bet. However, this effect is reversed and weaker when x = 0.5 because this is a small amount of money. The effect is most clearly seen when N = 5, where X = 1 has a much higher proportion of ‘yes’ responses than N = 10. When X = 1 or 2, as number of people increase, the proportion of yes decrease.

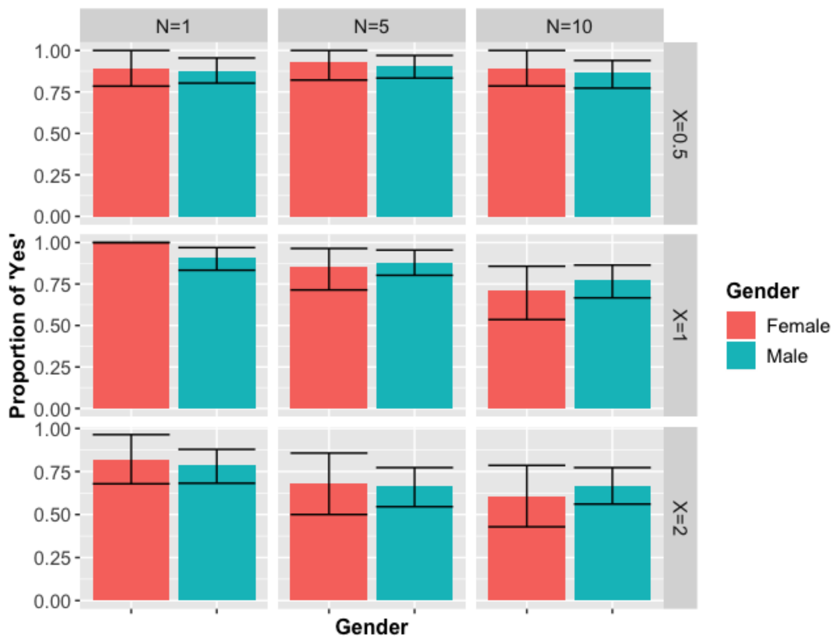

Figure 6: Proportion of agreement for cost per person (X) and number of other people(N), as grouped by gender, with bootstrapped 95% confidence interval

When we grouped by gender (figure 6), the results look similar and we cannot get any useful information. Gender may not be a significant predictor based on this figure but we still need to confirm in the discussion part.

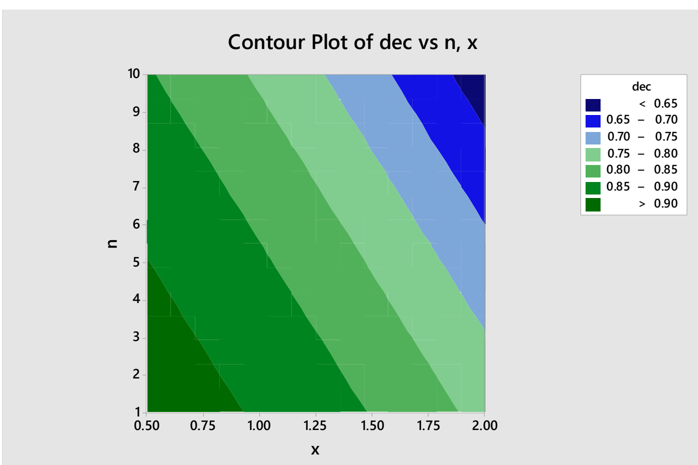

Figure 7: contour plot of decision verses N and X.

From the contour plot, we can see that X and N both have negative effect of decision. When N and X become smaller, the probability of decision increase.

Qualitative Analysis

We ask people the reason for their decisions on two, extreme sample questions in the fourth experiment:

"You have a choice to either earn $100 at a cost of $1 to 10 other people each, or to earn nothing but cost those ten people nothing. Do you choose to earn $100?"

"You have a choice to either earn $1 at a cost of $100 to 10 other people each, or to earn nothing but cost those ten people nothing. Do you choose to earn $1?"

These questions are included to help people understand that our survey depends on moral reasoning about money and to help them understand the rules of game. After each question, we asked participants for their reasoning.

The most frequent reason given by workers for their 'yes' decisions is that they actually need money. We cannot judge their answers because we know that hourly payment for MTurk workers is only $10 so we can easily understand why most people choose 'yes' in more than half of the questions. Only 3 of the 97 workers in Experiment 4 considered these questions in a moral manner. Their comments are: 1. We should not take money from others, no matter how much we want. 2. I have enough money. 3. I do not want to take money from others. From their comments, we can see that a few people really compare their personal gain and another group loss. Many of them only concern about individual rationality.

Quantitative Analysis

In the first three experiments, for the logistic regression model, the full model with N, X, Y, two way interactions and a three-way interaction was fitted first. The R output is shown as below:

Coefficient:				
	Estimates	Std. Error	Z Value	Pr(>IzI)
Intercept	2.312	0.195	11.838	< 2e-16
N	-0.238	0.050	-4.733	2.21e-06
X	-0.666	0.095	-6.994	2.68e-12
Y	-0.230	0.044	-5.257	1.47e-07
N:X	0.142	0.051	2.79	0.005257
N:Y	0.071	0.014	4.975	6.51e-07
N:X	0.156	0.023	6.685	2.31e-11
N:X:Y	-0.05	0.015	-3.464	0.000532

Table 2: Full logistic regression model.

Although every term in the full model is significant, it is very hard to interpret the main effect of N, X and Y with the presents of all the interaction terms. Therefore, a main effect model was then fitted.

Coefficient:				
	Estimates	Std. Error	Z Value	Pr(>IzI)
Intercept	1.33900	0.13617	9.834	< 2e-16
N	-0.02979	0.01893	-1.574	0.11555
X	-0.15290	0.04473	-3.418	0.00063
Y	0.08137	0.01716	4.742	2.11e-06

Table 3: Main effect logistic regression model

Based on the R output, without the interaction terms, N is not statistically significant, so we drop the factor of N, which brings us to our final logistic model selection:

Coefficient:	Estimates	Std. Error	Z Value	Pr(>IzI)
Intercept	1.20011	0.09340	12.698	<2e-16
X	-0.12995	0.03925	-3.049	0.000875
Y	0.07465	0.01646	4.503	5.9e-06

Table 4: Reduced multiple logistic regression model

The final model selected is: $\log(\frac{\pi}{1-\pi})$ = 1.20 – 0.13X + 0.075Y

AIC measures the quality of model. While losing all the interaction terms, the AIC for the final model did not change much. However, the final model is much easier to interpret compared to the full model. For the purpose of this project, the final model is ideal for the interpretation of the relationship we were looking for.

For general linear fixed model, the full model was first fit. And turns out the interaction terms are not significant. Then interaction terms were removed one at a time and the model was refitted. The final model with the lowest AIC is presented below.

Formula: Dec ~ I(Y^0.25) + I(X^0.75) + X:Y + (1| Subject)				
AIC	BIC	logLik	Deviance	Df.resid
1657	1685	-823.5	1647	1982

Table 5: Goodness of fit in GLMM model

Best model: $\log(\frac{\pi}{1-\pi})$ = 0.647 + $2.75Y^{0.25} - 0.45X^{0.75}$ + 0.078Y*X + (1| Subject)

Based on the output, N is not significant. However, the interaction between X and Y is statistically significant. This result is somewhat similar to what we got from logistic regression that X is negative correlated with the response and Y is positively correlated with the response.

In order to find the better model, we chose Pearson Goodness-of-fit statistic(chi-squared) to check how well the observed data correspond to each model with different indexes of Y and X. After we checked the X^2 for each model by fitting different exponents, the best model right now was the mixed model with Y's exponent equal to 0.25 and X's exponent equal to 0.75. The X^2 in this model is 990.8813 which has the smallest value compared with other models. In addition, the mixed model has smaller AIC value compare with the logistics regression model.

For the fourth experiment, our goal is to determine whether N also has a significant effect on the decision because only N is not significant in the previous three experiments. We set Y to a constant due to the fact that Y has the smallest p-value before. We also have to convert the predicting variable to factor in classification problems. After we got the clean data, the first step is to randomize the order of the data to prevent the order from affecting the composition of the training and test sets or the performance of the model. We split the clean dataset of responses from Experiment 4 into 80% training data (676) and 20% testing data (170). Our goal is to build the training model and then predict on the testing data.

We also fit the full-model logistic regression first and use the backward stepwise approach to remove the less significant feature.

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	3.36412	0.51319	6.555	5.55e-11
n	-0.12028	0.07010	-1.716	0.08619.
x	-1.00476	0.32535	-3.088	0.00201
n:x	0.02478	0.04572	0.542	0.58781

Table 6: Full logistic regression model

We can see that the interaction term and n are not significant because their p-values are more than 0.05. This is the same as previous results, so we used backward stepwise again to see the difference of different models.

	Degree of freedom	Deviance	AIC
-n:x	1	584.45	590.45
None		584.15	592.15
- n	1	593.48	597.48
- x	1	611.69	615.69

Table 7: Backward stepwise regression

This table proves that no interaction terms appear in our model. AIC is 590.45 which is the smallest of all models. Then we fit the model with no interactions of n and x.

Coefficients:	Estimates	Std. Error	Z Value	Pr(>IzI)
Intercept	3.14961	0.31414	10.026	< 2e-16 ***
N	-0.08576	0.02877	-2.981	0.00287 **
X	-0.85426	0.16647	-5.132	2.87e-07

Table 8: Multiple logistic regression with no interactions

These results are better than previous experiments because all variables prove to be significant, however, the R square is only 5.9%. The accuracy of final model is 82.3%.

The final model selected is: $\log(\frac{\pi}{1-\pi}) = 3.1496 – 0.086N – 0.854X$

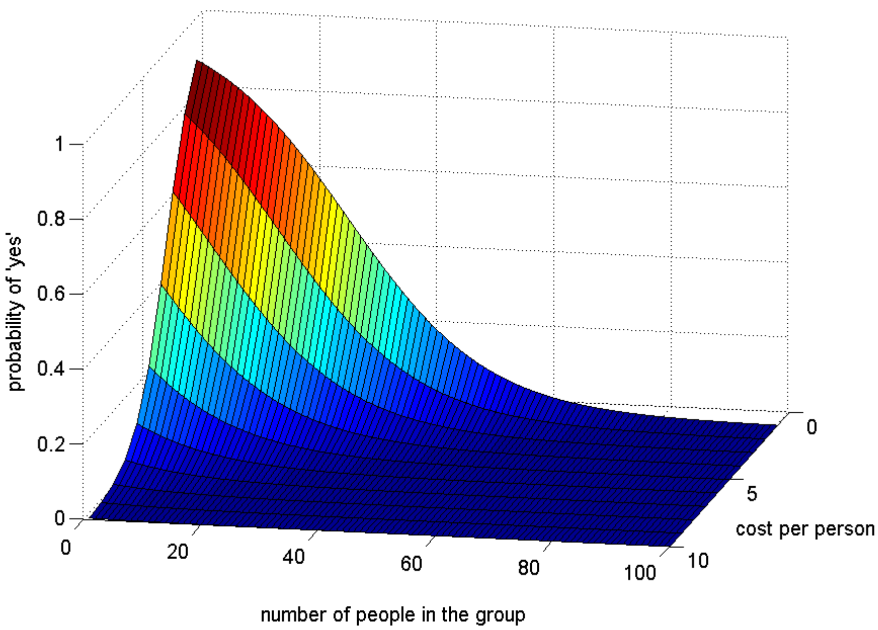

Figure 9: 3-D curve between two predictors and probability.

Based on the regression model, we draw the 3-D curve between two predictors and the probability of yes. This 3-D curve better give us a better view of how two predictors affect the probability of people saying 'yes'. we want to test how factor n will affect the probability of decision. We pick two cases: n=5, x=2 and n=10, x=1. The probability when is 0.735 when n=5, x=2 and 0.8094 when n=10, x=1. This can prove that n can improve the probability of saying yes but has less effect compared with other two predictors because the p-value of N is relative larger but less than 0.05 which means still within the 95% confidence interval. When number of people in the group increase, or cost per person increase, the proportion of 'yes' decision decrease.

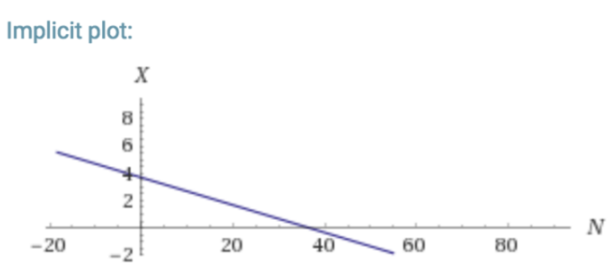

Figure 8: relationship between X and N about 50% ‘yes’

Next step is to evaluate the model in a wide range of values based on the empirical data. We set the threshold to 0.5 which means that if probability with more than 50% is considered to be true and otherwise false. The equation is :3.1496 – 0.086N – 0.854X = 0 = log (0.5/(1-0.5). From the implicit plot, we can see that the area behind the line in the first quadrant is the case of decision ‘yes’. The intercept to the y axis is 4 and intercept to the x axis is 38, so this means that if we compete with less than 40 people and the cost per person is less than $4, we would like to say yes. In other cases, we would prefer to say no out of that decision boundary.

Then we still need to test the quadratic effect so we fit the general linear mixed model. This time we still hold other variables fixed including the random effect. The coefficients estimates are below:

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	7.8668	1.0597	7.424	1.14e-13
n^0.25	-1.7975	0.4743	-3.790	0.000151
x^0.75	-1.9465	0.3368	-5.779	7.52e-09

Table 11: GLMM model with optimal exponents

Mixed model: $\log(\frac{\pi}{1-\pi}) = 7.866 - 1.7975N^{0.25} - 1.9465X^{0.75}$ + (1|subject)

We did the same steps again like previous experiments. The best model is that N’s exponent equal to 0.25 and X’s exponent equal to 0.75 with AIC is 472.9. The R squared of model is 44.3% which is much higher than logistic regression model.

Moreover, I have tried other prediction models like Naive Bayes model to make predictions and get the accuracy. We got the accuracy from the Naïve Bayes model first.

usekernel	Accuracy	Kappa
FALSE	0.8168155	0
TRUE	0.8168155	0

Table 9: Accuracy selected from the optimal model.

The highest accuracy for TRUE and FALSE are the same (0.8168) in Naive Bayes model. The final values used for the model were fL = 0, usekernel = FALSE and adjust = 1. The accuracy is similar to the logistic regression model.

Afterwards we use LDA and QDA model to make prediction and test the discrimination ability of logistic regression model by using the testing data. We firstly choose a cut point of 0.5 which means observations with a fitted probability above 0.5 is positive and below 0.5 is negative. For this threshold, we can estimate the sensitivity by the proportion of observations with Y=1 which have a predicted probability above 0.5, and similarly we can estimate specificity by the proportion of Y=0 with a predicted probability at or below 0.5.

The sensitivity is defined as the probability of the model predicting an observation as 'positive' given that in truth (Y=1). In simple words, the sensitivity is the proportion of positive observations which is classified as such by the model. In our case, we calculated the misclassification probability is 0.2294118. The sensitivity is equal to 1 and specificity is 0 so all predictions are true from the confusion matrix table.

We can conclude that LDA model is the same as baseline model (always say ‘yes’) and we cannot use the predictive methods in our experiments. The reason is that the probability of saying ‘yes’ in all conditions are more than 50%.

	Predicted: Yes	Predicted: No
Actual: Yes	169	0
Actual: No	31	0

Table 10: Confusion matrix

Moreover, we construct a ROC curve to further test the discrimination ability of model related to sensitivity and specificity. A Receiver Operating Characteristic Curve (ROC) is a common measurement for summarizing classifier performance over a range of tradeoffs between true positive and false positive error rates. ROC curve is a plot of sensitivity (the ability of the model to predict an event correctly) versus 1-specificity for the possible cut-off classification probability values $\pi 0$(Sweets, 1988).

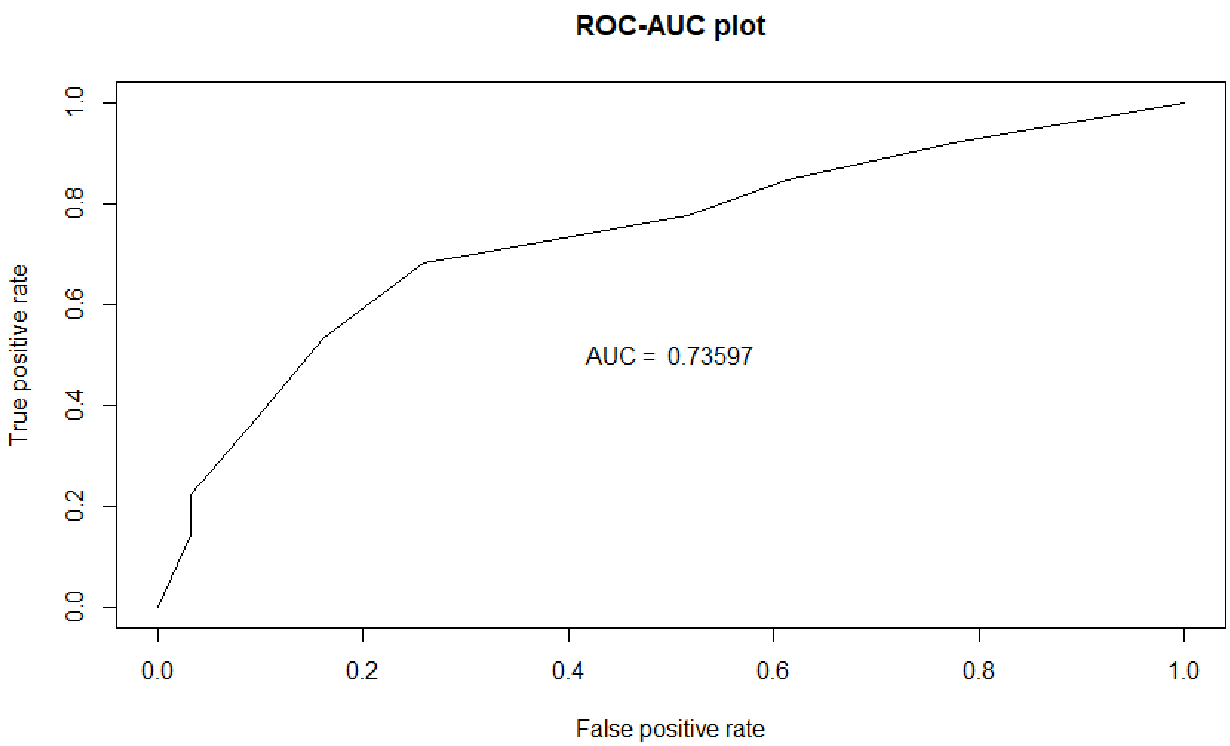

Figure 10: ROC-AUC curve

The accuracy of a test is measured by the area under the ROC curve. An area of 1 represents a perfect test, while an area of .5 represents a worthless test. The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test. The area under our curve is 0.736 and the best threshold we find is 0.7729929.

Based on the results from the ROC curve, I believe these prediction models are not quite useful in our project because probabilities of 'yes' we observed from each question are all more than 50% as mentioned before. Moreover, we only have two classes (yes and no) to predict, so usually two groups cannot be discriminated perfectly.

Dataset	Method	AIC	R squared	Chi-square

1st, 2nd & 3rd experiment	LG	2069.8	1.42%	2000.36
	GLMM	1671.5	6.78%	990.88
4th experiment	LG	590.45	5.9%	853.414
	GLMM	472.9	44.3%	510.909

Table 12: Performance comparison

From table 12, we can observe that AIC and chi-square values are smaller in the mixed model compared with logistic regression model which means mixed model is better. However, R squared values are not high in all cases. The problem is that our case is the classification problem, so R squared cannot represent the variability of response data. Furthermore, I also try some more complex algorithm to improve the accuracy of model, like decision tree and random forest. The accuracy does not improve a lot but the model is much more complicated.

In our problem, due to the size of datasets, GLMM model is more flexible to predict similar datasets because the variability is larger. Hence, we choose the generalized linear mixed model as our final model for this project.

Finally, I have to mention that we used 10-folder cross validation in the training data to improve the improve the performance because the size of our dataset is relative small. The accuracy increases from 80.3% to 81.7%. However, we find that cross validation is also not necessary in our case because this method relies on the prediction of model, so it cannot improve the accuracy of model. It is not meaning to give the accuracy since all models' accuracy under all conditions are more than 50% as mentioned before.

Chapter 6

Discussion

Chapter Overview

Whereas the results and findings section in Chapter 5 reported, the discussion section is mostly focused on limitations and improvements. In the final section, we also explored whether other factors that affect the decision, like year of birth and gender.

Limitations, weakness and potential improvement

The basic limitation is the lack of our sample size. As mentioned in Chapter 3, we need to avoid the case that we may put some all ‘yes’ answers into the testing data. Each experiment only has about 100 participants and a high proportion of ‘yes’ answers. To prevent the imbalance bias, we can allow to sample some data by multiple times. Another method is that we can duplicate someone’s answers who have more ‘no’s to better achieve the balance.

Based on first three experiments, our model was not as closely fit to the data as we had originally anticipated. The results on figure 1 and 2 partially verified our previous hypothesis that all three predictors should affect the proportion of decision, but N has different performance when N increase from 1 to 5 and 5 to 10, and Y affects a lot no matter of N. The main problem might be that bonus for each round is too small (2 cent), participants will not get any penalty or feel guilty when they say ‘yes’. We could introduce a risk factor like insurance into the model in the further classroom study, which will increase the risk of penalty when participants keep on answering yes. We are hoping with penalty for “yes” answers will push participant act more rational and give them some incentive to make ethical decisions. Another reason is that participants do not have any communication during the survey, so most of them would not consider it a real game.

Although we gradually improve our experimental design for the last experiment to make the game look more real and we try other algorithm, the accuracy of model is still not good enough. There are two major problems: relative small range of predictors and low income for MTurk workers in the Amazon platform. We can add a live chat room at the beginning of survey for MTurk workers to help them have some discussions. Each individual would enter a chat room to have a chance to talk to other participants but they all do not know the identity of individuals. This setting is only constructed to build a real gamble that make people convinced that they are participating in this game with other workers.

Age/Gender

In addition to modeling the characteristics of decision-making scenarios, we also explored how other demographic variables (e.g. gender, age, and country of origin) may affect the decisions. Firstly, I add gender as an independent variable too. We coded gender as a dummy, binary variable. While we found correlations between other variables and gender are all lower than 5%, the binary coding may have skewed the results. The sign of coefficient for the gender variable is positive, however, the p-values for gender are more than 0.90. This demonstrates that gender factor is not significant, so we can ignore it.

We also assessed if age plays a significant role in the decisions made by the subjects. After I do the same step as the gender, p-values for age is only significant at 95% confidence interval in the fourth experiment, however, the AIC just decreased 2.4 compared with the best logistic regression model. Therefore, we can get the conclusion that these two factors have slightly significant effect on the results.

Chapter 7

Conclusion

The experiment sought to gauge the acceptability of monetary loss of strangers for one's own gain. The background research of moral decision-making and group discounting lead the researchers to construct the experiment with emphasis on the threshold at which individuals decide to harm other individuals versus a group.

We checked the p-values of all the experiments and found that variable Y and X play an important role in the decision-making process. The effect of variable N (number of people) on the decision is less significant to affect people's decisions. We find that when keeping monetary gain (Y) constant, people are more likely to make ethical decisions when loss per person (X) goes up. On the other hand, while keeping the loss per victim (X) constant, people are less likely to make unethical decisions when monetary gain (Y) goes down.

Moreover, we only have three option values for all three predictors. Even though we may use our optimal model to predict similar datasets, but the accuracy will decrease a lot and the present model cannot predict a larger range of n and x since we only have three values for each predictor. Our model is not a good classifier and cannot predict other group of human data.

Since we keep total cost (N*X) constant in the first experiment, but this will lead to the problem that these two variables are collinear since we only have two variables to analysis each time. Thus, our experiments only confirm the individual rationality which means people are maximizing individual benefits because more than half of participants always choose 'yes' no matter in which condition. If we enlarge the range of three variables, the difference between personal gain and group loss will be larger and that is hard to conduct because we need more budget to increase the bonus.

Future Work

Even though we design many small details to make the experiment look like a real game, we still need to make more progress on the experiment design. My next objective is to test the group rationality because N and X are only collinear in the first experiment. The next experiment should still use the updated experimental design but have more comparison between Y (personal gain) and N*X (group loss). My hypothesis is that If Y is more than N*X, people would choose yes and otherwise choose no.

To improve our results, we intend to conduct more experiments with the university students. We could have a real gamble instead of simulations of games so that participants will not cheat and care more about money loss. The obvious advantage is that students can have interaction with others during the game. They can bet real money into each round when they compete with others. When they in reality lose some money, they would think more of questions morally. In summary, my aim is to further scientific understanding of the morality of monetary decisions by investigating how people think in terms of benefits to themselves and cost to others, and, in particular, cumulative cost across individuals.

The hope for future research in this field is to continue to study the relationship between participants' likelihoods to agree to harm others, potential profit, number of those harmed, and the severity of financial harm to others. In addition to studying behavior, new research regarding how the participants rationalize their decisions could provide beneficial in understanding the cognitive processes associated with moral decision-making.

References

Abdullah, A., Peter K., Yarrow D., David R., Alex P. (2018). Turkers of the World Unite: Multilevel In-Group Bias Amongst Crowdworkers On Amazon Mechanical Turk.

Amir, A., Kogut, T., & Bereby-Meyer, Y. (2016). Careful cheating: people cheat groups rather than individuals. Frontiers in psychology, 7.

Baron, J. (1995). Blind justice: Fairness to groups and the do-no-harm principle. Journal of Behavioral Decision Making, 8(2), 71–83.

Brief, A. P., Buttram, R. T., & Dukerich, J. M. (2001). Collective corruption in the corporate world: Toward a process model. Groups at work: Theory and research, 471, 499.

Gino, F., Shu, L. L., & Bazerman, M. H. (2010). Nameless+ harmless= blame- less: When seemingly irrelevant factors influence judgment of (un) ethical behavior. Organiza- tional Behavior and Human Decision Processes, 111(2), 93–101.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. Neuron, 44(2), 389–400.

Heekeren, H. R., Wartenburger, I., Schmidt, H., Schwintowski, H.-P., & Villringer, A. (2003). An fmri study of simple ethical decision-making. Neuroreport, 14(9), 1215– 1219.

Kahneman, D. (1979). Prospect theory: An analysis of decisions under risk. Econometrica, 47, 278.

Kugler, T., Kausel, E. E., & Kocher, M. G. (2012). Are groups more rational than individuals? a review of interactive decision making in groups. Wiley Interdisciplinary Reviews: Cognitive Science, 3(4), 471–482.

Shallow, C., Iliev, R., & Medin, D. (2011). Trolley problems in context. Judgment and Decision Making, 6(7), 593.

Shekar, P. (2018). Regression models in R. Retrieved from https://datascienceplus.com/how-to-perform-logistic-regression-lda-qda-in-r/

Soni, D. (2018). Introduction to Naive Bayes Classification. Retrieved from https://towardsdatascience.com/introduction-to-naive-bayes-classification-4cffabb1ae54

Trevino, L. K. & Youngblood, S. A. (1990). Bad apples in bad barrels: A causal analysis of ethical decision-making behavior. Journal of Applied psychology, 75(4), 378.

Appendix A

First Three Experiments Survey

General questions

1. How old are you?
2. What is your gender?
3. In which country did you grow up or spend the majority of your youth?

Rule of game

You are involved in a gambling game with 50 other random Amazon Turkers. In each round, you will be presented with a piece of information. You will then use this information to decide whether or not to participate in the game.

You will gain 0.01 dollars for completing each round and you can gain additional bonus depending on your answers. However, your decision in each round could potentially add to your bonus by taking away the bonus from other Turkers. Keep in mind, others will be playing a forced gamble on you too. The other players will also be presented with information that could also affect the gamble and your gain.

Your own reward will be revealed to you at the end of this game. However, your final bonus will be tallied after all the other Turkers compete in this game and will be applied within 36 hours of the HIT concluding.

Main question:
Read this gamble carefully. You're playing a game with a group of other Turkers. Your responses in each round affect other Turkers' bonuses, and their responses affect your bonus.

You are in your group of fifty turkers online right now, working this game like you do. You are betting $0.50 of your money in this round.
The other participants have also bet $0.50 each in this round.

You can now take $0.05 of the bets away from 10 other turkers and give it to the bank. Are you willing to tax N number of people that would lose X dollars for your N dollars gain?
You are guaranteed a 1 cent bonus for answering this question, regardless of your choice.

Appendix B

Fourth Experiment Survey

General questions

1. What's your username?
2. How old are you?
3. What is your gender?
4. In which country did you grow up or spend the majority of your youth?

Example questions

1. You are not playing with other workers at this time.
You have a choice to either earn $100 at a cost of $1 to ten other people, or to earn nothing but cost those ten people nothing. Would you choose to earn $100?
2. What was the reason for your decision in the example question?

3. Example question: You are not playing with other workers at this time.
You have a choice to either earn $1 at a cost of $100 to ten other people, or to earn nothing but cost those ten people nothing. Would you choose to earn $1?
4. What was the reason for your decision in the example question?

Rule of game

You are involved in a gambling game with other Amazon Turkers. In each round, you will be presented with an offer. You will then decide whether or not to accept the offer.

You will gain 50 cents for completing the survey and you can participate in a raffle to win an additional bonus depending on your answers. However, your decision in each round could potentially add to your bonus by taking away the bonus from other Turkers. Keep in mind, others will be playing a forced gamble on you too. The other players will also be presented with information that could also affect the gamble and your bonus.

You will start the game with a raffle ticket for $5 and then you can win more raffle tickets depending on your answers. Each raffle ticket has a fixed 1% chance of winning. The value of the prizes on your tickets can be affected by other Turkers' decisions. The raffle will take place within 36 hours of the HIT concluding with winners announced at https://bit.ly/2E1VJHM. Bonuses will be awarded to the winners at that time.

Main Question:

Read this gamble carefully. This is a real gamble and other workers will also be gambling against you. Your responses in each round affect other Turkers' bonuses, and their responses affect your bonus.

You are in your group of ten turkers online right now, the workers' usernames are: Tommysion,AholicJerry, ChrisChilli, Jameshu,KeepupAholic, BabyJamie, Jimmyqin, SnowBlue, StudioSparkling, Smugaboo.

You are betting $0.50 off one of your raffle tickets. The other participants have also bet $0.50 each in this round.

In this gamble, if you accept their bet, you can take $0.50 of the bets away from 1 other turker and give it to the bank. Are you willing to accept their bet for a raffle ticket of $5 prize?