Essays in Experimental Economics

Submitted by

Xiaoyu Zhou

for the degree of Doctor of Philosophy

of the

Royal Holloway, University of London

2016

Declaration

I, Xiaoyu Zhou, hereby declare that Chapters 1 and 2 of the thesis are my own research work. Chapters 3 and 4 comprise of collaborative research. Chapter 3 was conducted in collaboration with Dr. Michael Naef and Dr. Christoph Eisenegger. My contribution to this paper consists discussing and refining the research question, designing and coding the experiment, analyzing the data, and writing up the paper. Chapter 4 was conducted in collaboration with Dr. Michael Naef and Dr. Bjoern Hartig. My contribution to this paper consists discussing and refining the research question, designing and coding the experiment, running the experiment, analyzing the data, and writing up the paper.

Signed ..... (Xiaoyu Zhou)
Date:

Abstract

This thesis consists of three essays which utilize experimental methods to address different questions in the field of economics. The second chapter is related to the field of cultural economics, it investigates whether rice cultivation that practiced hundreds of years ago give rise to a cooperative social norm that has profound and long lasting effects on individual's behavior in an incentivized and strategic setting. The second chapter investigates whether testosterone leads to reputation enhancing prosocial behavior in the context of charitable donations, which is related to the field of neuroeconomics. In the forth and final chapter, we aim to disentangle the effect of two distinct aspects of reputation concern, namely, pursuit of honor and avoidance of shame, on charitable behavior.

Acknowledgements

I would like to express my deepest gratitude to my supervisor Michael Naef. Thank you for being patient and supportive, thank you for your invaluable advices and counsel, and thank you for orienting me to become a competent experimental economists. I am not there yet, but you have shown me how to proceed. I am in great debts to Bjoern Hartig, for spending time with me talking about research ideas, for the precious comments and suggestions for all my projects, and for helping me in my z-Tree coding. I am also grateful to Sotiris Georganas, for his advise and support for my visit to UPF. It was a wonderful year and I learned a lot.

Special thanks to my fellow PhD students who made these years more enjoyable. Especially Theo, for some times we were the only two experimentalist in the department. Special thanks to your helpful advices for my third paper. Neil, someday you will become an experimentalist.

Thanks to Prof. Fengtao Zhao and Prof. Lezhong Wang, for the support in arranging the experiments for the second chapter. Thanks to all the teachers and students who assisted me during the journey.

To my parents, thank you for the unconditional support and encouragement. You are the best parents, I love you.

To my wife, Yang, for your company for all these years. It is you who make London feel like home to me, Sharing all my highs and lows and never loose faith in me even during my hardest times.

For Kexin and Aixin

Contents

List of Figures **7**

List of Tables **9**

1 Introduction **10**

2 Rice Farming and the Emergence of Cooperative Behavior **12**

2.1 Introduction 12
2.2 Experimental Design 17
2.2.1 Province selection 17
2.2.2 Subject Recruitment 17
2.2.3 Experimental Games 19
2.2.4 Experimental Procedure 22
2.3 Experimental Results 23
2.3.1 Regression Analyses 30
2.4 Alternative Explanations 37
2.4.1 Rice Farming and Social Preferences 37
2.4.2 Differences other than Rice Cultivation 39
2.4.3 Self selection into rice and non-rice regions 41
2.5 Discussion 42
2.6 Conclusion 43

3 Does testosterone foster reputation concerns? **48**

3.1 Introduction 48
3.2 Experimental Design 51
3.2.1 The real effort task 52
3.2.2 Experimental Conditions and Treatments 53

3.2.3 Experiment Procedure 55
3.3 Hypotheses 56
3.4 Results 59
3.5 Discussion 64
3.6 Conclusion 66

4 The Virtue of Honor and the Power of Shame – an Experiment on Charitable Donations 67
4.1 Introduction 67
4.2 Experimental Design 71
4.2.1 The real effort task 71
4.2.2 Experimental Conditions and Treatments 72
4.2.3 Honoring and Shaming ceremony 74
4.2.4 Experimental Procedure 75
4.3 Results 76
4.3.1 Identification 76
4.3.2 The Effect of Honor and Shame 77
4.3.3 The Effect of Reputation Concerns on Different Characteristics. 78
4.3.4 Robustness Checks 84
4.4 Conclusion 85

Appendices 88
A Appendix to Chapter 2 89
A.1 Relative Welfare 89
A.2 Additional Regression Analyses 91
A.3 Experimental Instructions 96
B Appendix to Chapter 3 116
A Appendix to Chapter 4 123
A.1 Response to Different Pay Levels 123
A.2 Experimental Instructions 124

Bibliography 132

List of Figures

2.1 Average contribution in each period in PGG and PGG with Punishment 25
2.2 Punishment Behavior 28
2.3 Behavior in the Ultimatum Game and Dictator Game 38

3.1 The real effort task – Find the 'T's 54
3.2 The real effort task – The different symbols 54
3.3 The Sequence of Events in Each Session 56
3.4 Histogram of the treatment effect in honor and shame treatments . . 60
3.5 2D:4D Ratio (Digit Ratio) 62
3.6 2D:4D Ratio (Digit Ratio) – Distance from the median 62
3.7 Social Desirability 63
3.8 Behavioral Inhibition System (BIS) and Behavioral Activation System (BAS) 64

4.1 Information screen before each period starts 73
4.2 Treatment effect by median split of reputation concern scale 79
4.3 Treatment effect by median split of Fear of negative evaluation scale . 81
4.4 Treatment effect by median split of Ability measure 82
4.5 Regressions about the differential effects of reputation concerns on subjects' characteristics 83
A.1 Evolution of relative welfare 89
A.1 How Subjects Respond to Different Pay Levels 123

List of Tables

2.1 Payoff matrix of the Stag Hunt Game 22
2.2 Random Effects Panel Regression about the different contribution patterns in rice and non-rice provinces 26
2.3 Regressions regarding the Probability of Punishment and the Intensity of Punishment Conditional on Punishing 31
2.4 Random Effects Panel Regression on how subjects respond to punishment 32
2.5 Description of Control Variables 34
2.6 Random Effects Panel Regressions regarding the contribution in the PGGs. 35
2.7 Tobit Regressions regarding amount of punishment assigned to free-riders and cooperators. 36
2.8 Regressions regarding the Probability of Punishment and the Intensity of Punishment Conditional on Punishing using Municipality level Rice statistics 44
2.9 Random Effects Panel Regressions about reciprocity in the PGGs . . 45
2.10 Random Effects Panel Regressions Comparing contribution in the PGG with punishment between the two rice provinces and the two non-rice provinces. 46
2.11 Tobit Regressions Comparing punishment behavior between the two rice provinces and the two non-rice provinces. 47

3.1 Summary of Experimental Treatments 55
3.2 Regressions for the interaction effects 65

4.1 Summary of Experimental Treatments 74
4.2 Average amount of money made in each condition 77

4.3 Random effects panel regression 86
4.4 Random effects panel regression - Between subjects data structure . . 87
A.1 Tobit Regressions Regarding the Effect of the Percentage of Rice Paddy fields on Punishment Behavior among Rice Farming Regions. . 91
A.2 Random Effects Panel Regressions regarding the contribution in the PGGs. 92
A.3 Tobit Regressions regarding amount of punishment assigned to free-riders and cooperators. 93
A.4 Comparing the Probability of Punishment and the Intensity of Punishment Conditional on Punishing between subjects from rice provinces 94
A.5 Comparing the Probability of Punishment and the Intensity of Punishment Conditional on Punishing between subjects from non-rice provinces 95

Chapter 1

Introduction

This thesis consists of three essays which utilize experimental methods to address different questions in the field of economics. Chapters 3 and 4 are related as they both investigate people's behavior in the context of charitable donations. Chapter 2 aims to explore the origins of the observed differences in the level of cooperativeness across societies and countries, which contributes to the literature of culture economics and hence is unrelated to the other two chapters.

Chapter 2 aims to explore the origins of observed differences in cooperativeness across societies and cultures. In particular, the paper shows that rice cultivation practiced hundreds of years ago gives rise to a more cooperative social norm that affects contemporary decision making in a incentivized and strategic setting. The rationale behind this finding is that unlike cultivating other crops, such as wheat or corn, farming rice historically required extensive cooperation among farmers. Therefore, hundreds of years of rice farming might lead to the creation of a cooperative social norm that affects people living in that society, and the social norm transmits fairly unchanged from generation to generation. To test this hypothesis, I travelled to four typical rice and non-rice provinces in China and recruited a total of 524 local *university* students as subjects. The results indicate that rice subjects contribute more than their non-rice counterparts in the Public Goods Game with and without punishment, with the effect being a lot more pronounced in the former. Further analyses reveal a significant difference in frequency, though not in magnitude of punishment. There is no difference in how the two groups react to punishment. It follows that the different level of cooperativeness observed in the punishment treatments is a direct effect of the significant differences in frequency of punishment.

Furthermore, as there is no difference between the two groups in the ultimatum and in the dictator game, the interpretation of the results is that rice cultivation does not make people more cooperative per se. Instead, it is more likely the case that people in rice regions understand the nature of public goods differently than their non-rice counterparts. I attribute the differences to a cultural norm resulting from a history of farming that affects the whole population living in that society.

Chapter 3 investigates whether testosterone leads to prosocial behavior in the context of charitable donations. Researches in the field of psychology and neuroscience have established that testosterone's main role in human social interaction is to promote dominance behavior, that is, actions to maintain high status in situations where individuals face challenges from others. In this paper, we take a step forward by investigating whether testosterone leads to reputation enhancing behavior in absence of status threats. In a real effort donation experiment, a single dose of 0.5 mg of testosterone or placebo was applied sublingually in a randomized and double-blind fashion. The results indicate that there is no significant behavioral differences between the testosterone and placebo group, which is inconsistent with our hypotheses. We argue that three factors, namely, all female subjects, real effort task too enjoyable, and lack of observations are the main reasons of why we find null results.

In the forth and final chapter, we aim to disentangle the effect of two distinct aspects of reputation concern on charitable behavior. It is well established that reputation concern is one important motivation underlying charitable behavior. However, less is know whether it is the pursuit of honor, avoidance of shame, or the combination of both that drive people's behavior. These two motives are separated in a carefully designed donation experiment. The results indicate that both the desire to gain prestige and the urge to avoid social disapproval are important motives in explaining giving behavior.

Chapter 2

Rice Farming and the Emergence of Cooperative Behavior

2.1 Introduction

Studies have documented that people's level of cooperativeness differs across countries and societies.[1] However, the reason that these differences exist remains an open question. Understanding the factors that make people more cooperative is important, as the level of cooperation is associated with economic and political outcomes such as economic development (Knack and Keefer, 1997) as well as norms of civic cooperation and rules of law (Herrmann et al., 2008).

In this paper, we show that traditional agricultural activity practiced hundreds of years ago gives rise to a cooperative social norm that persists over generations and has profound influence on people's behavior in an incentivized and strategic setting. The agricultural activity refers to rice cultivation. Two features of rice farming distinguish itself from other crops, such as wheat and corn. First, rice grows on standing water instead of dry land, and farmers in a village traditionally share the same water reserves. This resulted in the need to cooperate in the management of the common resource and in the maintenance of the irrigation system. In addition, rice farming requires a large amount of labour. Farmers first need to grow rice

[1] See Herrmann et al. (2008) and Henrich et al. (2006) for differences in punish behavior in the Public Goods Game (PGG), Gächter et al. (2010) and Gächter and Herrmann (2009) for differences in cooperation in the PGG, and Jackson and Xing (2014) for different behaviors in the coordination game.

seedlings in separate lands and then transplant rice seedlings into paddy fields. This necessitated a large amount of labour. Indeed, agricultural anthropologists conclude that the amount of labour required in rice cultivation is at least twice as the number needed in farming wheat (Buck, 1935). Therefore, families with a only a few labor forces will not be able to survive if they rely on rice farming (Hsiao-Tung and Chih-i, 1945). To solve the shortage of labour supply during farming and harvesting seasons, farmers in rice villages form cooperative labour exchanges. Farmers also summon their relatives who live in neighbor villages to deal with the labour shortage issue. In sum, farming rice historically requires extensive cooperation among farmers, and the hypothesis is that centuries of rice farming leads to a cooperative social norm that is intergenerational transmittable.

To test this hypothesis, we travelled to typical rice and non-rice provinces in China and conducted lab experiments at local universities. We believe that China is a perfect testbed for the rice theory. Agriculture has been the most important industry throughout China's history. During China's imperial era, most of the population was farmers.[2] Even in 2011, about 34.8% of China's population was employed in the agriculture sector (World Bank). More importantly, China also has a long history of rice cultivation (Fan, 2007).

Following Talhelm et al. (2014), the categorization of rice and non-rice regions is based on the proportion of cultivated land devoted to rice paddy fields. We use the earliest available data from the National Bureau of Statistics website, because we do not want the farming statistics be affected by recent advances in technologies, but rather to reflect, as closely as possible, the historical farming choice. Indeed, the two rice provinces in our sample have been prominent rice production provinces since Song Dynasty (960 – 1279) (Fan, 2007).

We recruited local, Han Chinese, first year university students based on their Hukou. Hukou is a household registration system employed in China. The policy requires that individuals must register the Hukou at their city of residence and they can only register their Hukou at one city. By local students we mean that their Hukou was registered at the province of experiment. For example, if the experiment was conducted in Hebei, which is a typical non-rice province, subjects with Hebei Hukou were recruited.[3]

[2]See "A Brief History of China's Economy" edited by Fudan University and Shanghai University of Finance and Economics (1982)

[3]Please see to the experimental section for a detail discussion of the recruitment criteria and

To measure the difference of cooperativeness between rice and non-rice subjects, we implemented the Public Goods Game (PGG) with and without punishment. Moreover, in order to gain a deeper insight of the rice and non-rice difference, we also asked subjects to participate in a dictator game (DG), an ultimatum game (UG), and a coordination game. We also implemented a priming treatment in half of the sessions in which we simply informed subjects that they were from the same province (please see the experimental design section for details). Results suggest that priming has no effect on subject's behavior in the experimental tasks and we therefore pool the data for analyses.

There are several important features of our experimental design worth mentioning. First and foremost, subjects were not farmers but university students who had minimal farming experiences.[4] Therefore, the present study goes beyond merely testing the effect of rice farming on individual's cooperativeness. Instead, we aim to investigate whether hundreds of years of rice farming is capable of creating a cooperative social norm that affects everyone in the society and that transmits fairly unchanged from generation to generation. Second, we travelled to the selected rice and non-rice provinces and conducted the experiments locally. We believe that the sample in our study is more representative and has less selection issues than studies that conduct experiments in one Beijing University and recruit students from all over China in that university.[5] The reason is that Beijing universities set small quotas for students from provinces other than Beijing, therefore, non-Beijing students who intend to study in Beijing face stronger competition. Consequently, the majority of high school graduates choose local universities. Third, by recruiting Han Chinese University students, a number of potential confounds are controlled by design, such as educational background, language, culture, and political institution. And last, we have a relatively rich data set that allows us to link individuals to administrative data such as their Hukou type (Rural or Urban) and the local GDP measure.

We find that rice subjects contribute more than non-rice subjects in both the PGGs with and without punishment. In the no punishment condition, although rice subject's contribution is higher, the difference is relatively weak. Rice subjects also experience a stronger end game effect, as the contribution unravels to the same

their advantages.

[4]See the experiment design section for a discussion of why this is the case.

[5]A large proportion of subjects in Talhelm et al. (2014) were recruited in Beijing

level as non-rice subjects in the last period. To the contrary, the difference in contribution is substantial in the punishment condition. Rice subject's contribution is already significantly higher than non-rice in the first period. More importantly, the difference not only sustains but also enlarges towards the end. We also find that rice subjects are more likely to punish free-riders compared to their non-rice counterparts and there is no difference in the intensity of punishment conditional on punishing. Additionally, the two groups do not differ in how they respond to punishment. It follows that the different levels of cooperativeness observed in the punishment condition is a direct consequence of the significant differences in punishment behavior. Furthermore, as there is no difference between the two groups in the ultimatum game and in the dictator game, we interpret from our results that rice cultivation does not make people more cooperative per se. Instead, it is more likely the case that people in rice regions understand the nature of public goods differently than their non-rice counterparts. And last, we run a series of robustness checks and show that the conclusions are less likely driven by self-selection and omitted variable bias.

To the best of our knowledge, this is the first paper that shows agriculture activities that practiced hundreds of years ago have a long lasting and profound effect on individual's level of cooperativeness and punishment behavior in a incentivized and strategic setting. The paper most related to ours is Talhelm et al. (2014), who found that people from rice and wheat areas in China have different thinking styles, measured by the Triad task.[6] However, unlike Talhelm et al. (2014), we focus our attention on economically relevant behavior in incentivized and strategic situations. We also administrated the Triad task and find no difference between the rice and non-rice subjects in our sample. More importantly, our results are not attenuated by controlling for the thinking style. This suggests that our results are mediated from a different channel other than thinking styles. The methodological difference is that we used validated and incentivized games from the experimental economics literature, while they used non-incentivized questionnaires.

This research contributes to the emerging literature that aims to explore the origins of the observed differences in people's preferences across societies. Alesina

[6]We use the rice and non-rice terminology instead of rice and wheat. This is because in additional to wheat, there are other types of crops that require less cooperation to farm, such as corn. Note that the categorization of rice and non-rice provinces is exactly the same as rice and wheat.

et al. (2013) investigates the origins of cross-cultural differences in norms about gender roles. They find that societies which historically practice plough agriculture has less equal gender norms in the work place. The rationale behind this observation is that plough requires considerable level of upper body strength, hence, men have a relative advantage to women in workplace in those societies. Galor and Özak (2016) find that participation in agricultural activities affects people's time preferences. In particular, the higher but delayed return feature of the agriculture sector makes people more future orientated. Nunn and Wantchekon (2011) find that the level of mistrust within Africa origins from the transatlantic and Indian Ocean slave trades.

The present paper differs from the aforementioned papers in three important ways. First and foremost, we focus on explaining the origins of cross cultural differences in the level of cooperativeness, while they focus on gender norms, time preferences, and trust. Second, their measures mainly come from surveys, while in our data, subject's cooperativeness is elicited in an incentivized and strategic setting. And last, we believe that we have a more direct approach to show that our results are less likely driven by self-selection. In particular, we are able to identify the subjects in our sample who had migrated into or out of rice farming regions, and we show that our findings are not affected by the exclusion of these subjects.

There is also a small but emerging literature that utilize lab or lab-in-the field experiments to explore the observed differences in preferences across countries. Gneezy et al. (2016) and Leibbrandt et al. (2013) find that sea fishermen are more cooperative and less competitive than lake fishermen. They argue that this is because the difference in work place organization between the two groups: sea fishing requires intensive team work among crew members in order to survive in the sea, while lake fishing is usually an individual activity. However, they do not find any behavioral differences among women in the two societies who do not fish, and they conclude that "... suggestive evidence that norms of cooperation learnt at the workplace do not spread to other society members" (Gneezy et al., 2016, p. 2). This is in sharp contrast to our finding because subjects in our experiment are all university students who have minimal farming experience. In a related paper, Carpenter and Seki (2006) also find evidence supporting the idea that work place organization has profound influences on cooperation.

The remainder of the paper is structured as follows: Section 2.2 illustrates the experimental design. Non-parametric and regression results are presented in sec-

tion 2.3. A series of robustness checks are conducted in section 2.4. Section 2.5 discusses the findings, and section 2.6 concludes.

2.2 Experimental Design

The idea of the design is to compare subjects who have similar backgrounds but live in regions that differ sharply in the type of cultivation. The most important feature of our design is that the subjects were university students instead of farmers. Thus, the present paper aims to investigate whether rice farming is capable of creating a cooperative social norm that transmits fairly unchanged from generation to generation.

2.2.1 Province selection

We conducted lab experiments in four universities that are located in four provinces across China. By restricting the sample to Chinese subjects, a number of potential confounds are controlled by design, such as language, political institution, and other cross country cultural differences.

Since we are investigating the intergenerational transmission of social norm created by rice farming hundreds of years ago, the categorization of rice and non-rice provinces should be based on the type of crops farmers historically cultivate. For this purpose, we use the earliest available cultivation data (1996) on the Bureau of Statistics Web site.[7] Following Talhelm et al. (2014), a province is classified as rice if more than half of its cultivated land is devoted to rice paddy field. The two rice provinces we choose are Hunan and Zhejiang province, which devote more than 78% of the cultivated lands to paddy fields. Also note that these two provinces has been prominent rice farming provinces since Song Dynasty (Fan, 2007). The percentage is less than 2.5% in the non-rice provinces: Hebei and Shandong.

2.2.2 Subject Recruitment

We recruited local, Han, first year university students. We believe that our subjects have minimal farming experiences. The reasons are twofold. Firstly, about 40% of

[7] http://www.stats.gov.cn

the subjects in our sample hold an urban Hukou which implies that they do not have the legal right to possess farming lands.[8] Second, since high schools are generally located in larger cities, commuting between school and home might be too costly, in terms of time and money, for students who have a rural Hukou. Therefore, they might choose to live on campus during school days and hence do not have much farming experience. One might argue that they can help their families during weekends or holidays. However, we doubt this limited time of farming experience is able to modify their behavior. In addition, rice cultivation nowadays probably requires less cooperation due to advances in technology, therefore, even if subjects spend sufficient time in rice farming, the norms of cooperation in the field might not be strong enough to alter their behavior. Thus, if there is any difference in cooperation between rice and non-rice subjects, it is unlikely due to their personal farming experiences but the intergeneration transmission of social noms related to the crops that used to be predominantly cultivated in their regions.

We recruited local university students for three important reasons. First, this reduces the chance of recruiting subjects from less typical provinces. For example, in Sichuan and Jiangsu province, the percentage of farming land devoted to rice and non-rice crops are very close between each other. Second, since reciprocity is an important motivation in the PGGs (Fischbacher et al., 2001), behavioral norms might be hard to emerge if rice and non-rice subjects interact. And third, having a local Hukou suggests that the subject has more likely been living in the area for a long time and hence is more affected by the social norm.[9]

We chose Han Chinese because Wen et al. (2004) discovers that Han Chinese have the same culture origin. We also recruited first year students because they had just graduated from high school. Chinese high schools have a busy schedule: students stayed in classroom more than seven hours every day and study similar materials. Therefore, subjects had similar experience prior to university. Moreover, first year students are free from the indoctrination effect of their field of study.[10]

[8] According to China's Hukou policy, people with Rural Hukou are entitled with farming land.

[9] When analyzing the data, we find that a few subject's Hukou was not from the province of experiment. In order to utilized as many observations as possible, we drop subjects whose Hukou was from a different farming province. For example, if the experiment was conducted in rice provinces, we drop subjects who had Hukou from non-rice farming provinces. The results are the same if we drop subjects whose Hukou province was different from the experiment province. Please note that in the latter case, more observations are dropped.

[10] There were 9 subjects who were not first year students. Including or excluding them do not

The recruitment process was the following. A list of qualified students was provided by each university, and we randomly drew subjects from the list. Selected subjects were then contacted by the administrative staffs of each university. We provided a script about how to recruit the students. We emphasized that it was a economic study, they would receive money payments to compensate for their time, their decisions in the study would be anonymous and would not affect their records related to university in any way, and most importantly, participation was *not* compulsory.

2.2.3 Experimental Games

To compare the level of cooperation between rice and non-rice areas, we conducted a repeated public goods game (PGG) with and without punishment. We believe that the situation farmers encounters everyday is very similar to the situation in the PGG: each farmer has the incentive to free-rider on other farmers during planting or harvesting seasons. However, the society reaches the most efficient outcome if all farmers exert maximum effort.

Subjects first played eight periods under the no punishment condition followed by eight periods of punishment condition. They knew that there would be another game after the no punishment condition, but they were not informed about its content until the no punishment condition was completed.

In the no punishment condition, subjects are randomly divided into groups of four and the group composition is fixed throughout the eight periods. In each period, each subject has an endowment of 20 points and is asked to decide how many points to contribute to a group account (the remaining points are allocated to their individual account). The total points in the group account are multiplied by 1.6 and then evenly distributed among all group members. In particular, each subject face the following payoff function:

$$u_i = (20 - c_i) + (1.6 * \sum_{j=1}^{4} c_j) \ / \ 4$$

in which u_i is i's payoff, c_i is i's contribution to the group account, and $\sum_{j=1}^{4} c_j$ is the sum of contribution made by all group members.

affect our results and hence we include them in the analyses.

Note that the contributor only gains 0.4 points for each point contributed to the group account. Therefore, contributing nothing always give subjects the highest material payoff regardless of other group member's contribution. On the other hand, each point contributed to the group account increases the payoff of the whole group by 1.6 points, and hence the group level payoff is highest if all group members contribute 20 points. In the latter case, each subject earns 32 points, which is higher than the self interested outcome, 20 points.

After all subjects make their decisions, the amount of contribution of each subject, their earning from the group account, and their total earning in the current period are shown on their computer screen. The contribution of each group member is randomly displayed on the computer screen in each period, therefore, subjects cannot associate contribution with a particular group member.[11] Subjects need to press the 'CONTINUE' button to proceed to the next round.

After the no punishment condition, subjects randomly regrouped and play eight periods of the punishment condition. The first part of the punishment condition is the same as the no punishment condition: each subject has an endowment of 20 points and need to decide how many points to contribute to a group account. After this decision is made, subjects are informed about their earnings from the first stage and are asked to proceed to punishment stage, in which subjects can assign punishment tokens to other group members. In the punishment stage, other group members' contributions in the present period are displayed on subject's computer screen. Based on this information, subjects can assign punishment tokens, which are restricted between zero and ten inclusive, to other group members. Each punishment token costs one point to the punisher and reduces the earnings of the punished subject by three points.[12] Next is the information display stage in which subjects' final earnings are shown. They are informed about their earnings in the first stage, total punishment tokens received and total punishment tokens assigned to others in the punishment stage, and their final earnings. Please note that subjects only know the total punishment tokens received but not who assigned the punishment.

[11] Other group members' contribution are displayed because subjects are also asked to play the PGG with punishment, in which case group members' contributions must be revealed. We intend to make the design of the two games as close as possible.

[12] Subjects are informed that their earning can only be reduced to zero no matter how many punishment tokens they receive. However, negative earning is possible if ones earning is reduced to zero due to receiving too many punishment tokens and she also assigns punishment tokens to others. This design is also used in Herrmann et al. (2008).

In order to gain a deeper insight of the rice and non-rice difference, we also asked subjects to participate in a dictator game (DG), an ultimatum game (UG), and a coordination game (Stag Hunt).

In the dictator game, subjects are randomly assigned to the role of proposers or responders. Proposers have to decide how to divide a total of 60 points between themselves and a randomly matched, anonymous responder. When the task starts, responders are asked to state how many points they expect to receive. Please note that responders' answers will not affect the outcome of the dictator game, and the proposers are not informed about this.[13]

Subjects' roles in the ultimatum game are the same as in the dictator game, but they are randomly regrouped. The difference between the ultimatum game and the dictator game is that in the former, responders have the power to reject or accept offers made by proposers. We employed a minimal acceptable offer (MAO) method. When proposers are making offers, responders simultaneously state their minimal acceptable amount. If the offer made by the proposer is lower than the minimal acceptable amount, the allocation is automatically rejected, in which case both of them earn nothing. If the offer is larger or equal to the minimal acceptable amount, the proposal is automatically accepted, in which case both of them receive the amount according the division. Subject's role in the DG and UG was fixed because this design allows us to investigate whether rice and non-rice subjects have different level of strategic consideration. Since the responder can reject an offer in the UG while has no influence in the DG, strategic individuals should offer nothing in the DG and offer a higher amount in the UG. Therefore, the difference in the offer amount in the UG and DG is a measure of strategic behavior. Results suggest that there is not difference between rice and non-rice in this regard.

The stag hunt game is a two-player simultaneous move coordination game. The payoff matrix is presented in table 2.1. Subjects can choose between hunting a stag or a hare. Stag is harder to catch but more valuable. Both players need to choose the same action to make the hunt successful. If they mis-coordinate, the one who choose stag will fail and hence obtain the lowest payoff. Hare, on the other hand, is easy to catch and is therefore a safe choice: it yields a payoff of 22 points regardless of other player's action.

[13]This is accomplished by displaying the information on responders computer screen after the DG starts.

Table 2.1: Payoff matrix of the Stag Hunt Game

	Stag (@)	Hare (#)
Stag (@)	(30, 30)	(12, 22)
Hare (#)	(22, 12)	(22, 22)

Subjects' risk attitudes are elicited using the Holt and Laury type lotteries (Holt and Laury, 2002). They are informed that this task is not incentivized. We also implement the triad task, which is the main dependent variable in Talhelm et al. (2014). The Triad task is designed to measure people's thinking styles. The questionnaire presents subjects with a list of three objects, and subjects are asked to choose the two items that they think are more related to each other. For example, one of the questions is panda, banana, and monkey. Panda and monkey is an analytic choice because they are both animals. On the other hand, monkey and banana is a holistic choice since monkey eats banana. Rice and non-rice subjects in our sample do not differ in the level of thinking styles.

Since the literature in psychology suggests that collective societies value group membership, we conduct a priming treatment in half of the sessions. The procedure is simple. In Hebei for example, after all subjects arrived in the lab and were waiting for instructions, the experimenter stated: please note that all of you are from Hebei province.[14] In the other sessions, subjects are not informed about this information. We find that priming has no effect on subject's behavior, we therefore pool the data from priming and no-priming sessions in the analyses.

2.2.4 Experimental Procedure

After all subjects arrived in the lab and prior to getting any instructions of the study, they were asked to sign a formal consent. Participants knew that each session consists several parts, but they did not know the content of the future parts until the corresponding instructions were provided.

The order of the experimental tasks was organized as follows. First, subjects were asked to fill out a questionnaire, which consisted the nonincentivized lottery

[14]In some sessions, few subjects stated that they were not from the local province. In this case, the experimenter explained that they cannot participate in the experiment and were free to leave. Of course, the show-up fee was paid to them.

task and the triad task. Second, they played the DG, UG, Stag Hunt game, the PGG no punishment condition, and the PGG punishment condition, in that order. After they competed a post experimental questionnaire, they received their payment and were free to leave.

The reason we set the order of the games as previously described is that we intend to avoid the outcome of one game to affect subject's behavior in the subsequent games. The DG, UG and Stag Hunt are the first three games because we can easily withhold the outcome of these games until the very end of each session. This is not possible in the PGGs. In the no punishment, the amount of contribution of each group member and each subject's earning is shown after each period. In the punishment condition, the act of punishing or getting punished might influence subject's behavior, we therefore allocate the punishment condition to the last. In sum, the order of the games aims to minimize the externality of each game on subject's subsequent behavior.

One of the five games was randomly selected for payment. If the PGGs were chosen, the experimenter would draw one period out of the eight periods. Subject's earning were exchanged to Chinese Yuan at the rate: 1 points = 0.5 Yuan (about 8 US Cents).

The experiment was conducted between Oct 2015 and Jan 2016 in China. There were a total of 524 subjects. 116 subjects for Hebei province, 156 subjects in Shandong province, 128 subjects in Hunan province and 124 subjects in Zhejiang province. We ran 6 sessions per province. All the sessions were conducted on Saturdays and Sundays because student's schedule was busy during weekdays. Each session lasted about 2 hours. Subjects earned on average 30 Yuan (about 5 US dollars), including a 15 Yuan show-up fee. Subjects' earnings were similar to China's minimal hourly wage.[15]

2.3 Experimental Results

If the tradition of rice farming has resulted in the creation of a more cooperative social norm and this norm is transmittable from generation to generation, one should

[15]See Appendix A for the experimental instructions. Please note that the instructions are in English, as they were used for a pilot session conducted in Royal Holloway, University of London. Moreover, we abandoned the sliding bar in the DG, UG, and PGGs for the sessions in China. Input boxes were used instead. The Chinese version is available upon request.

observe that subjects from rice farming provinces contribute more than subjects from non-rice provinces in the public goods games. This is exactly what we find. The following result summarizes the findings in the PGG without punishment.

Result 1. *In the PGG without punishment, rice subject's contribution is higher than non-rice subject's contribution. However, the difference does not sustain in the last period.*

Support for result 1 is presented in figure 2.1. Figure 2.1 illustrates the evolution of average contribution over time. In the no punishment condition (periods 1 - 8), the average contribution over the eight periods is 10.70 and 9.63 points for rice and non-rice regions respectively. The difference is marginally significant (Mann-Whitney U test, each group as an independent observation: $p = 0.0657$).[16]

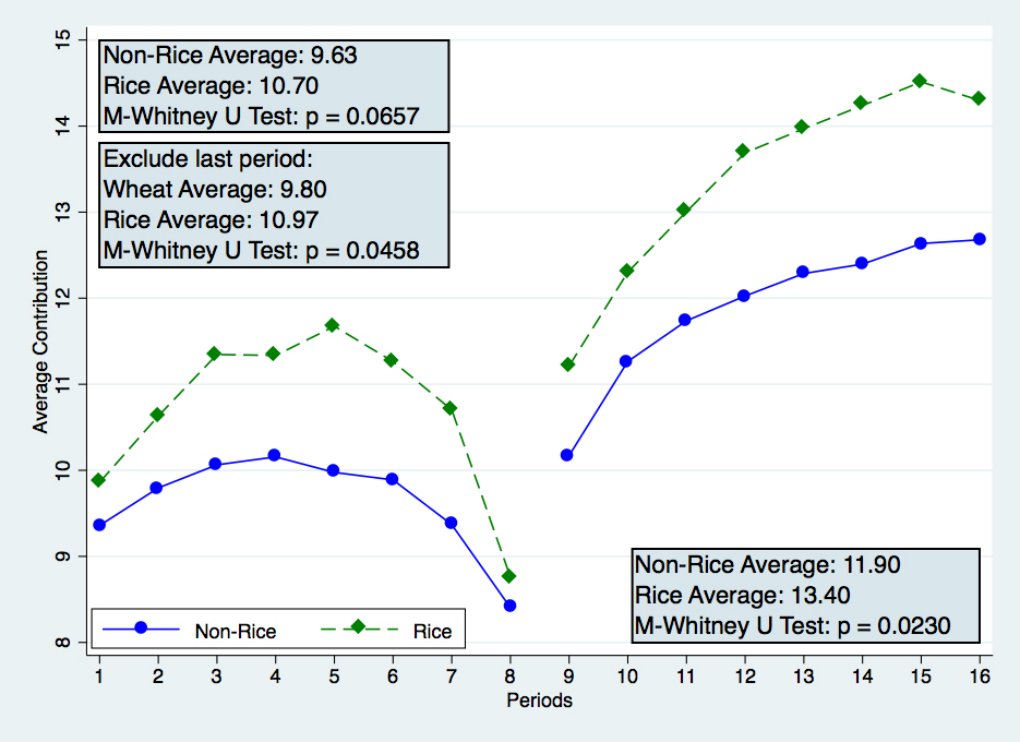
Although the difference in average contribution between rice and non-rice subjects in the no punishment condition is only marginally significant, the difference in the pattern of contribution over time is dramatic. Rice and non-rice subjects start at similar level of contribution in the first period (Mann-whitney U test, each group as an independent observation: $p = 0.0937$). Difference begins to emerge over time. Contribution of non-rice subjects increases modestly over interaction. In contrast, rice subjects manage to increase contribution dramatically. However, rice subjects also experience a stronger end game effect, as the contribution unravels to the same level as non-rice subjects in the last period (Mann-whitney U test, each group as an independent observation: $p > 0.66$).[17]

The different contribution pattern between rice and non-rice is also confirmed in the random effects panel regression showed in table 2.2. Both the Period and Period Squared are highly significant, suggesting that the contribution pattern of non-rice subjects exhibits a inverted-U shape. The interaction terms between the Rice dummy and the period terms are also signifiant, which implies that rice subject's contribution pattern has more curvature.

[16] We use group level average and conduct statistical tests based on group level to control for within group dependency. However, the group level average is not exactly the same as the individual level average due to the fact that some subjects are dropped from the analyses. Please note that the results do not change if we conduct Bootstrapped ttest on group level clusters.

[17] The difference in contribution between rice and non-rice in the PGG without punishment becomes significant if the last period is excluded because of the strong end game effect (Mann-Whitney U test, each group as an independent observation: $p = 0.0458$).

Figure 2.1: Average contribution in each period in PGG and PGG with Punishment



Notes: The x-axis is period number and y-axis is the average contribution. Periods 1-8 are public goods game without punishment. After period 8, subjects randomly regroup and play another eights rounds of public goods game with punishment (periods 9-16).

The stronger end game effect of rice subjects compared to non-rice subjects suggests that rice farming might not make people more cooperative per se, instead, it might be the case that people from rice areas understand the public goods situation differently than non-rice subjects. In other words, they understand that it is profitable to contribute to the public good so long as there are future interactions.

If it is indeed the case that rice subjects are more familiar with the public good situation, one should expect the difference in contribution between rice and non-rice is more substantial in the punishment condition, since punishment has proven to be highly effective in fostering cooperation (Fehr and Gächter, 2000; Masclet et al., 2003). Subject's behavior in the PGG with punishment is summarized in result 2.

Result 2. *In PGG with punishment, rice subject's contribution is significantly higher than non-rice subject's contribution. The difference is already significant in the first period and enlarges towards the end.*

The evidence of result 2 is presented in period 9-16 of figure 2.1. The average

Table 2.2: Random Effects Panel Regression about the different contribution patterns in rice and non-rice provinces

	(1)	
Rice	-0.292	(0.535)
Period	0.827***	(0.208)
Rice x period	0.761**	(0.337)
Period squared	-0.105***	(0.0215)
Rice x period squared	-0.0810**	(0.0342)
Constant	8.588***	(0.348)
Observations	4136	

Notes: The regression is a random effects panel regression. The dependent variable is the contribution in the no punishment condition. There are 4136 observations, which implies 517 subjects are included in the analysis. As mentioned previously, 7 out of 524 subjects are dropped from the analyses because they come from a province that the main type of crop is different from the province of experiment. Cluster Standard errors (PGG group level) are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

contribution over the eight periods is 11.90 points for non-rice and 13.40 points for rice; the difference is significant (Mann-Whitney U test, each group as an independent observation: p = 0.02). Note that the difference in contribution is already significant in the first period (Mann-Whitney U test, each group as an independent observation: p = 0.02). In the presence of punishment opportunities, the difference not only sustains but also enlarges towards the end, which is in stark contrast to the no-punishment condition.

One important feature of figure 2.1 is that for both rice and non-rice subjects, the contribution increases sharply from the no-punishment condition to the punishment condition. We elaborate on this observation in more detail in the following.

Result 3. *The presence of punishment opportunity significantly increases contribution for both rice and non-rice subjects. Moreover, punishment has a slightly stronger effect for rice subjects.*

Non-rice subjects on average contribute 9.63 points in the no punishment condition. This number increases to 11.87 in the presence of punishment. The difference, 2.24 points, is highly significant. For rice subjects, the increment is 2.73 and is also highly significant (Wilcoxon sign-ranked tests: p < 0.01 for both rice and non-rice subjects).[18] The results also suggest that punishment has a stronger effect for rice subjects. The difference in contribution between the no-punishment and punishment

[18]Since subjects randomly regroup after the no-punishment condition, it is impossible to conduct paired tests based on the group level. Therefore, these two tests are based on individual level.

condition is larger for rice subjects, and it is marginally significant (Mann-Whitney U test: p = 0.0572).

In the following, we investigate subject's punishment behavior. We are interested in how subjects punish each group member instead of the *sum* of punishment points assigned to all group members in each period. We also distinguish between prosocial and anti-social punishment. Prosocial punishment is defined as assigning punishment points to subjects who contribute less than the punisher. We label this behavior pro-social punishment because the punisher is willing to sacrifice her own payoff to punish free-riders, and free-riders who receive punishment are more likely to increase their contribution in following periods (Fehr and Gächter, 2000; Gächter et al., 2008; Masclet et al., 2003; Nikiforakis, 2008). Therefore, pro-social punishment is beneficial to the whole group. Anti-social punishment is defined as punishing group members that contribute more than or equal to the punisher. This behavior is labelled anti-social punishment because the punished subject behaved more pro-socially than the punisher and this behavior detrimental to the group's payoff (Herrmann et al., 2008).

Distinguishing between pro-social and anti-social punishment is important because individuals have different punishment behavior towards free-riders and cooperators (Fehr and Gächter, 2000), and this is a common practice in the literature.[19].

Result 4a. *Rice subjects make significantly more pro-social punishments than non-rice subjects. There is no difference in antisocial punishment.*

Result 4b. *Further analyses of the data suggest that rice subjects are significantly more likely to punish free-riders, there is no difference in the intensity of punishment conditional on punishing free-riders.*

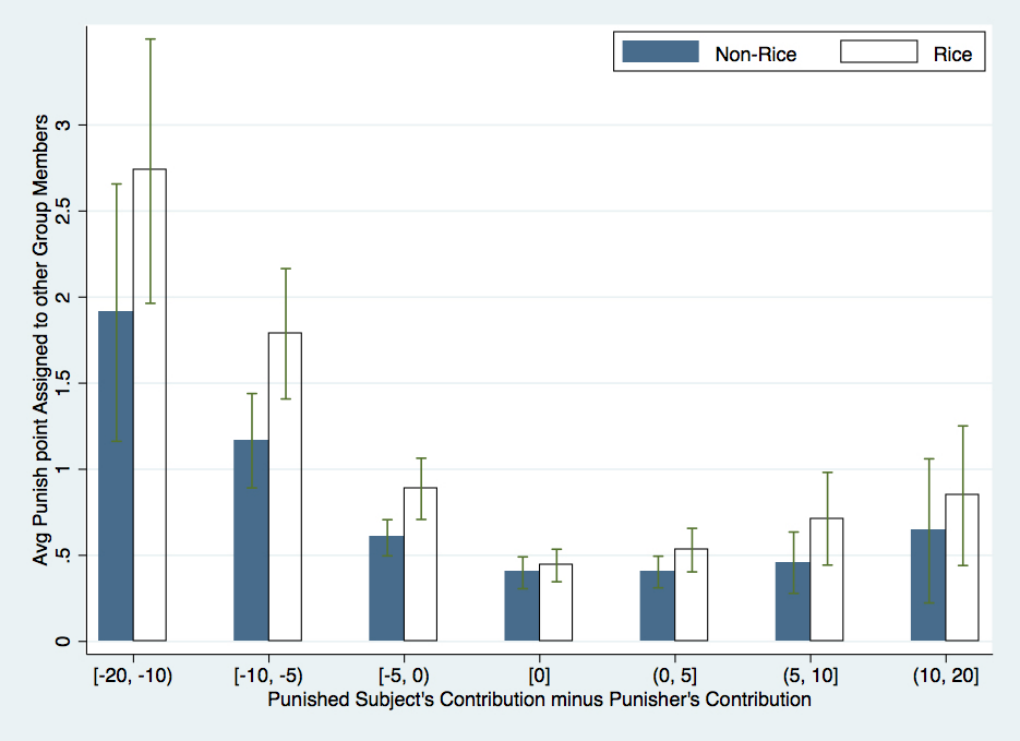
Evidence for Result 4.a is provided in figure 2.2. Figure 2.2 depicts the average punishment points assigned to others as a function of deviation from the punisher's contribution. For example, the [-20,-10) category implies that the punished subject contributes from 10 to 20 points *less* than the punisher. Therefore, all the categories to the left of [0] are pro-social punishment, and all the other categories are anti-social punishment. The results confirm that rice subjects assign significantly more punishment points to free-riders. The difference is significant for categories [-10, -5)

[19]See (Fehr and Gächter, 2000; Masclet et al., 2003; Denant-Boemont et al., 2007; Anderson and Putterman, 2006; Faillo et al., 2013; Herrmann et al., 2008)

and [-5, 0) (Mann-Whitney U test, $p < 0.01$ for [-10,-5) and $p = 0.02$ for [-5,0)) and is weakly significant for the [-20,-10) category (Mann-Whitney U test: $p = 0.08$). If we merge all the three free-riding categories into one category, the difference between rice and non-rice becomes highly significant (Mann-Whitney U test, $p < 0.01$). On the other hand, there is no significant difference between rice and non-rice areas in the anti-social categories (Mann-Whitney U test: $p > 0.22$ for all relevant categories: [0], (0, 5], (5, 10], and (10, 20]).[20]

The punishment pattern in figure 2.2 also suggests the necessity to separate prosocial and anti-social punishment. As shown in the figure, for pro-social punishment, the amount of punishment point assigned is positively associated with the size of free-riding. This association, however, is much weaker when the punisher contributes more than the punished subject.

Figure 2.2: Punishment Behavior



Notes: This figure shows how subjects punish those who contribute less, more, or equal to themselves. The x-axis shows the difference between ones own contribution and one of her group member's contribution. For example, [-20, -10) implies one of my group member's contribution is from 10 to 20 points *less* than my contribution. The y-axis is the average punishment point subjects assigned to each category.

[20] If we merge all the four positive deviation categories into one category, the difference between rice and non-rice becomes only weakly significant (Mann-Whitney U test, $p = 0.08$).

Additionally, we find that the percentage of cultivated land devoted to rice farming also predicts pro-social punishment among rice regions; there is no such relationship among non-rice regions. (see table A.1 in Appendix A.2. Columns 1-2 are for rice regions; columns 3-4 are for non-rice regions).

Evidence supporting Result 4.b is presented in table 2.3. Columns 1-3 of table 2.3 show that rice subjects are more likely to punish free-riders in comparison to their non-rice counterparts (we use the Random effects Linear Probability Model for the ease of presentation, the results are the same under the Random Effects Probit model). In column 1, the Rice dummy is positive and significant at 10% level (it is highly significant in the Random Effects Probit Model), confirming that rice subjects are more likely to punish free-riders. The Anti-Social Punishment variable is significantly negative, which means that non-rice subjects are significantly less likely to punish cooperators.

Following Brañas-Garza et al. (2014) we also controlled for the punisher's contribution and the punished subject's contribution (Column 2). We also controlled for the other two group member's average contribution. Without these controls, one cannot claim that rice subjects are more likely to punish free-riders, instead, it might simply due to the fact that the contribution difference between the punisher and punished subject is larger in rice areas. For example, suppose there are two groups, A and B. The punisher in both groups contributes 15 points. In group A, one group member contributes 5 and in group B, one group member contributes 12. In this case, the punisher in group A are more likely to punish the free-rider simply because the contribution difference is too large. Controlling for the punisher and punished subject's contribution allows one to compare the tendency to punish other group members while holding the contribution difference constant. The reason to include the other two group member's average contribution is similar.

Columns 4-6 of table 2.3 shows that conditional on punishing, the intensity of punishment does not differ between rice and non-rice subjects.[21] This is true for both anti-social punishment and pro-social punishment. In column 4, the variable Rice is not significant, suggesting that conditional on punishing free-riders, the amount of punishment points assigned to free-riders does not differ between rice and non-

[21] We used tobit model to analyze punishment points assigned to other groups members because the amount of punishment is restricted between zero and ten inclusive. Note that the results are the same if we use linear models.

rice. The interaction term between Rice and Anti-Social Punishment is also not significant. This implies that there is no difference in the intensity of anti-social punishment conditional on performing an anti-social punishment. The results are the same if we control for the contribution level of other group members (column 5).

Result 5. *There is no difference between rice and non-rice in how subjects respond to punishment.*

Table 2.4 demonstrates the results on how subjects respond to prosocial punishment (columns 1-2) and anti-social punishment (columns 3-4). The dependent variable is the change in contribution from period t to t+1 conditional on receiving punishment points in period t. The variable Punish Receive is the *total* number of punishment points received in period t. Also note that the definition of prosocial and anti-social punishment is slightly different from previous analyses. Here, prosocial punishment implies that the punished subject contributes less than other group member's average contribution. Anti-social punishment is defined analogously. This adjustment is crucial because it is impossible to separate the effect of each punisher's punishment points assigned to the punished subject.

Results in column 1 indicate that non-rice subjects significantly increase their contribution after receiving pro-social punishment. In particular, they increase 0.675 points for each pro-social punish point received. The interaction term "Punish Receive × Rice" is not significant, suggesting that there is no difference between rice and non-rice subjects in how to react to pro-social punishment. The result for antisocial punishment is shown in column 3. The "Punishment Received" is now significantly negative, which implies that non-rice subjects decrease their contribution after receiving antisocial punishment. The interaction term "Punish Receive × Rice" is not significant, which means rice and non-rice subjects also behave similarly in the response to anti-social punishment.

2.3.1 Regression Analyses

In the previous section, we show that rice subjects contribute more in the PGGs with and without punishment and make more prosocial punishment compared to non-rice subjects. All the results are either based on non-parametric tests comparing

Table 2.3: Regressions regarding the Probability of Punishment and the Intensity of Punishment Conditional on Punishing

	Probability of Punishment			Intensity of Punishment		
	(1)	(2)	(3)	(4)	(5)	(6)
Rice	0.0704*	0.0926**	0.0864**	0.453	0.431	0.308
	(0.0410)	(0.0402)	(0.0387)	(0.278)	(0.266)	(0.241)
Anti-Social Punishment	-0.138***	-0.0810***	-0.0807***	-0.139	0.327	0.257
	(0.0148)	(0.0189)	(0.0189)	(0.164)	(0.251)	(0.250)
Rice X Anti-Social Punishment	-0.0163	-0.00946	-0.00933	0.00559	0.0542	0.236
	(0.0246)	(0.0244)	(0.0244)	(0.342)	(0.353)	(0.333)
Punisher Contribution		0.0107***	0.0106***		0.0595	0.0392
		(0.00306)	(0.00307)		(0.0368)	(0.0311)
Punished Contribution		-0.0102***	-0.0102***		-0.0376**	-0.0399**
		(0.00199)	(0.00199)		(0.0190)	(0.0170)
Other Two Member Avg Contribution		-0.0187***	-0.0186***		-0.0246	-0.0305
		(0.00307)	(0.00308)		(0.0292)	(0.0258)
GDP per cap. (Municipality)			-0.000171			0.00276
			(0.00457)			(0.0275)
Holistic Thinking			-0.150**			0.299
			(0.0661)			(0.411)
From Rural (Dummy)			-0.000112			0.299
			(0.0339)			(0.197)
Relative Income			0.00231			0.135
			(0.0207)			(0.141)
Risk Attitude			0.0103			0.141*
			(0.00788)			(0.0775)
Male			0.0183			0.687***
			(0.0282)			(0.243)
Natural Science (Dummy)			-0.00252			0.222
			(0.0259)			(0.169)
Priming (Dummy)			0.0819**			-0.127
			(0.0357)			(0.215)
Single Child (Dummy)			-0.0161			-0.137
			(0.0319)			(0.247)
Collectivistic			0.000413			0.0702
			(0.0338)			(0.254)
Individualistic			0.0632*			0.396*
			(0.0349)			(0.216)
Trustworty (Belief)			-0.00920			0.0222
			(0.00810)			(0.0590)
Public Order (Belef)			0.00160			0.0195
			(0.00608)			(0.0507)
Period	-0.0233	0.0198	0.0206	0.103	0.148	0.0197
	(0.0458)	(0.0454)	(0.0453)	(0.293)	(0.289)	(0.283)
Period Squared	0.000555	-0.000961	-0.000985	-0.00600	-0.00758	-0.00204
	(0.00182)	(0.00180)	(0.00180)	(0.0117)	(0.0116)	(0.0112)
Constant	0.554**	0.445	0.323	1.736	1.220	-1.090
	(0.278)	(0.277)	(0.308)	(1.811)	(1.807)	(2.146)
Observations	7519	7519	7519	2379	2379	2379

Notes: Columns 1-3 are Random Effects Linear Probability models and the dependent variable is the probability of punishing Free-rider and Cooperators. Columns 4-6 are tobit regressions and the dependent variable is the amount of punishment points assigned to other group member who contributed more or less than the punisher. Cluster Standard errors (PGG group level) are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

Table 2.4: Random Effects Panel Regression on how subjects respond to punishment

	Pro-Social Punishment		Anti-Social Punishment	
	(1)	(2)	(3)	(4)
Punish Receive	0.675***	0.663***	-0.362***	-0.375***
	(0.110)	(0.107)	(0.137)	(0.138)
Punish Rec. X Rice	0.0287	0.0439	0.294	0.302
	(0.172)	(0.170)	(0.211)	(0.205)
Rice	0.0744	-0.0285	-0.125	-0.0678
	(0.290)	(0.291)	(0.164)	(0.166)
GDP per cap. (Municipality)		-0.00270		-0.0139
		(0.0254)		(0.0207)
Holistic Thinking		0.353		0.636**
		(0.412)		(0.263)
From Rural (Dummy)		-0.0236		0.0450
		(0.249)		(0.152)
Relative Income		-0.236*		0.0966
		(0.125)		(0.119)
Risk Attitude		-0.00918		-0.0220
		(0.0499)		(0.0588)
Male		0.493**		-0.0912
		(0.223)		(0.139)
Natural Science (Dummy)		0.488***		0.166
		(0.179)		(0.144)
Priming (Dummy)		0.0355		-0.00209
		(0.190)		(0.164)
Single Child (Dummy)		0.524**		-0.120
		(0.230)		(0.144)
Collectivistic		-0.0131		0.0478
		(0.177)		(0.185)
Individualistic		-0.0853		-0.169
		(0.208)		(0.156)
Trustworty (Belief)		0.0192		0.0805
		(0.0412)		(0.0531)
Public Order (Belef)		-0.000724		-0.00852
		(0.0398)		(0.0479)
Period	-0.951**	-0.907**	0.0393	0.0631
	(0.467)	(0.459)	(0.480)	(0.484)
Period Squared	0.0371*	0.0353*	-0.00672	-0.00779
	(0.0197)	(0.0194)	(0.0207)	(0.0208)
Constant	6.730**	6.296**	0.0443	-0.813
	(2.705)	(3.058)	(2.731)	(3.059)
Observations	1461	1461	1465	1465

Notes: the dependent variable is the change in contribution from t to t+1 conditional on receiving punishment points in period t. Punish receive is the total number of punishment points received in period t. Columns 1 and 2 investigates cases in which the punished subject contributes less than other group member's average contribution (response to prosocial punishment). Columns 3 and 4 investigates cases in which the punished subject contributes more than other group member's average contribution (response to anti-social punishment). Cluster Standard errors (PGG group level) are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

subjects from rice areas to subjects from non-rice areas, or regression analyses using the Rice dummy as the main independent variable of interest.

In the following, we use a finer level of rice statistic, which is the percentage of cultivated land devoted to rice paddy fields at the municipality level, to validate the tests conducted in the previous section. Note that in this specification, the variable of interest is a continuous variable (the Rice dummy used previously is based on the province level rice statistics. Regressions using the Rice dummy as independent variable are presented in the Appendix A.2 and the results are similar).[22] Moreover, we show that our results are robust to the inclusion of a set of control variables. The set of control variables and their descriptions are presented in table 2.5.[23]

Table 2.6 illustrates that the higher the percentage of cultivated land devoted to rice paddy fields, the higher the contribution in the PGGs, which is congruent with Result 1 and Result 2 (see table A.2 in the Appendix A.2 for the results using the Rice dummy variable). Columns 1 and 2 present the results for the no punishment condition. The variable of interest is positive and significant at the 10% level. It implies that a 10% increase in the percentage of cultivated land devoted to paddy field is associated with a 1.2 increase in contribution. This result is not attenuated by the inclusion of control variables (Column 2).

Subject's contribution in the punishment condition is shown in columns 3 and 4. The coefficient in column 3 suggest that a 10% increase in the percentage of cultivated land devoted to paddy field is associated with a 1.8 increase in contribution, and it is significant at the 5% level. This result is also robust to the inclusion of the control variables.

Table 2.7 shows that the percentage of paddy field is positively associated with pro-social punishment but not anti-social punishment, which confirms Result 3 (see

[22]Note that for both the dummy and continuous specification, we also ran the regressions using the Hierarchy Model under five levels: Province level, Municipality level, Session level, PGG group level, and Individual level. The results are the same.

[23]Three subjects are dropped in the regression analyses with control variables. One subject did not provide their Hukou place at the municipality level so we cannot match him to the rice statistic. One subject did not state whether he has a Rural or Urban Hukou. Moreover, these two subjects provided the incorrect student number so we cannot recover the information using administrative data. Another subject need to leave early, so she did not answer the social style questionnaire. In order to keep the number of observation consistent between regressions with and without control variables, we dropped these three observations in the regressions without control variables. Note that the results are not affected if we do not drop these subjects in the regressions without control variables.

Table 2.5: Description of Control Variables

Variable Name	Description
GDP per capita (Municipality)	GDP per capita at the municipality level. The latest data available (2014) from the province level Bureau of Statistics websites and matched with subject's Hukou place.
Holistic Thinking	Percentage of holistic choices in the Triad task. The main dependent variable in Talhelm et al. (2014).
Collectivism & Individualism	In the highly influential work, Hofstede (1980) proposed the questionnaire in order to compare societies based on different social styles. The core element of Individualism is that individuals are independent of each other. On the other hand, group membership is a central aspect of collectivism.
Priming (Dummy)	A dummy variable equals one if the data comes from the priming treatment (See the Experimental Design section for the description of the priming treatment).
From Rural (Dummy)	A dummy variable equals one if the subject Has a Rural Hukou, equals zero otherwise.
Relative Income	Relative income assesses subjects income level relative to their town of residence. There are four levels, "Way Above", "Above", "Same", "Below", and "Way Below".
Risk Attitude	The number of risk seeking choices in the non-incentivized Holt & Laury lottery task.
Single Child (Dummy)	A dummy variable equals one if the subject has no siblings, equals zero otherwise.
Natural Science (Dummy)	According to China's Education Policy, students in senior high school need to choose between two screams, the social-science-oriented area, which focuses on history, politics and geography and the natural-science-oriented area, which focuses on physics, chemistry and biology. In the National Higher Education Entrance Examination, the two streams have separate exam papers. Please note that students who choose the social-science-oriented also need to study physics, chemistry and biology, but to a lesser degree. This is also true for students in the natural-science-oriented.
Trustworthy	Measures to what extend subjects believe that people from the local province are trustworthy. Likert scale: 0 "Strongly Disagree" and 10 "Strongly Agree".
Public Order (Belief)	Measures to what extend subjects believe that people from the local province obey public order, for example, do not jump queues, do not spit, and do not shout in public spaces. Likert scale: 0 "Strongly Disagree" and 10 "Strongly Agree".

Table 2.6: Random Effects Panel Regressions regarding the contribution in the PGGs.

	PGG without Punishment		PGG with Punishment	
	(1)	(2)	(3)	(4)
Perc. Paddy Field (Municipality)	0.0125*	0.0121*	0.0180**	0.0166**
	(0.00726)	(0.00696)	(0.00785)	(0.00733)
GDP per cap. (Municipality)		-0.0321		0.0454
		(0.0596)		(0.0608)
Holistic Thinking		1.375*		0.374
		(0.772)		(0.735)
From Rural (Dummy)		-0.265		-0.392
		(0.412)		(0.413)
Relative Income		-0.566**		-0.296
		(0.281)		(0.238)
Risk Attitude		0.0727		0.160
		(0.117)		(0.106)
Male		2.042***		2.179***
		(0.462)		(0.370)
Natural Science (Dummy)		0.651*		1.311***
		(0.380)		(0.393)
Priming (Dummy)		-0.461		-0.148
		(0.542)		(0.569)
Single Child (Dummy)		-0.345		-0.0580
		(0.477)		(0.409)
Collectivistic		0.499		0.692
		(0.443)		(0.466)
Individualistic		-0.629		-0.433
		(0.428)		(0.390)
Trustworty (Belief)		0.0784		0.123
		(0.0900)		(0.105)
Public Order (Belef)		-0.0358		-0.0469
		(0.0919)		(0.0807)
Period	1.191***	1.191***	2.244***	2.244***
	(0.171)	(0.171)	(0.301)	(0.302)
Period sqaured	-0.144***	-0.144***	-0.0746***	-0.0746***
	(0.0173)	(0.0174)	(0.0115)	(0.0115)
Constant	7.926***	7.284***	-4.118**	-7.531**
	(0.429)	(2.552)	(1.772)	(3.077)
Observations	4112	4112	4112	4112

Notes: The dependent variable is contribution in the PGG without punishment (columns 1 and 2) and with punishment (columns 3 and 4). Perc. Paddy Field is the percentage of cultivated land devoted to paddy fields at the municipality level. Cluster Standard errors (PGG group level) are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

table A.3 in appendix A.2 for the results using the Rice dummy variable). Column 1 shows that a 10% increase in the percentage of cultivated land devoted to paddy field leads to 1.1 more punishment points assigned to free-riders. Post regression test suggests that Perc. of Paddy Field is not associated with anti-social punishment ($p > 0.27$). These results are not affected by the inclusion of the set of control variables (column 2).

Table 2.7: Tobit Regressions regarding amount of punishment assigned to free-riders and cooperators.

	(1)		(2)	
Perc. Paddy Field (Municipality)	0.0111**	(0.00470)	0.0115**	(0.00455)
Anti-Social Punishment	-1.442***	(0.264)	-0.871***	(0.289)
Perc Rice Paddy X Anti-Social Punishment	-0.00431	(0.00470)	-0.00347	(0.00491)
Punisher Contribution			0.0582	(0.0396)
Punished Contribution			-0.0561***	(0.0209)
Other Two Member Avg Contribution			-0.102***	(0.0363)
GDP per cap. (Municipality)			-0.0291	(0.0432)
Holistic Thinking			-1.189**	(0.492)
From Rural (Dummy)			0.0467	(0.273)
Relative Income			0.147	(0.199)
Risk Attitude			0.144	(0.0911)
Male			0.462	(0.321)
Natural Science (Dummy)			-0.0420	(0.251)
Priming (Dummy)			0.615*	(0.359)
Single Child (Dummy)			-0.332	(0.346)
Collectivistic			0.109	(0.369)
Individualistic			0.691*	(0.368)
Trustworty (Belief)			-0.0544	(0.0785)
Public Order (Belef)			-0.00268	(0.0613)
Period	-0.244	(0.451)	-0.00224	(0.448)
Period Squared	0.00416	(0.0181)	-0.00426	(0.0179)
Constant	0.949	(2.743)	-2.139	(3.100)
Observations	7519		7519	

Notes: the dependent variable is the punishment points assigned to group members who contributes less or more than the punisher. Perc. Paddy Field is the percentage of cultivated land devoted to paddy fields at the municipality level. Anti-Social punishment is a dummy variable which equals 1 if the punished subject contributes more than the punisher. Cluster Standard errors (PGG group level) are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

In table 2.3, we show that subjects from rice farming areas are more likely to punish free-riders and conditional on punishing, there is no difference in the intensity of punishment (Result 4.b). These findings are also not affected by the set of control variables (columns 3 and 6). We replicate this result using the Perc. Paddy Field as the independent variable. The findings are presented in table 2.8 and are consistent

with the regressions using the rice dummy variable (table 2.3). Columns 1-3 of table 2.8 show that the percentage of rice paddy field is positively associated with the probability of prosocial punishment. However, it does not predict the intensity of punishment conditional on punishing (columns 4-6). Note that in columns 4 and 5, the Perc. Paddy Field coefficient is weakly signifiant. However, it looses significance after including the set of control variables (column 6).

2.4 Alternative Explanations

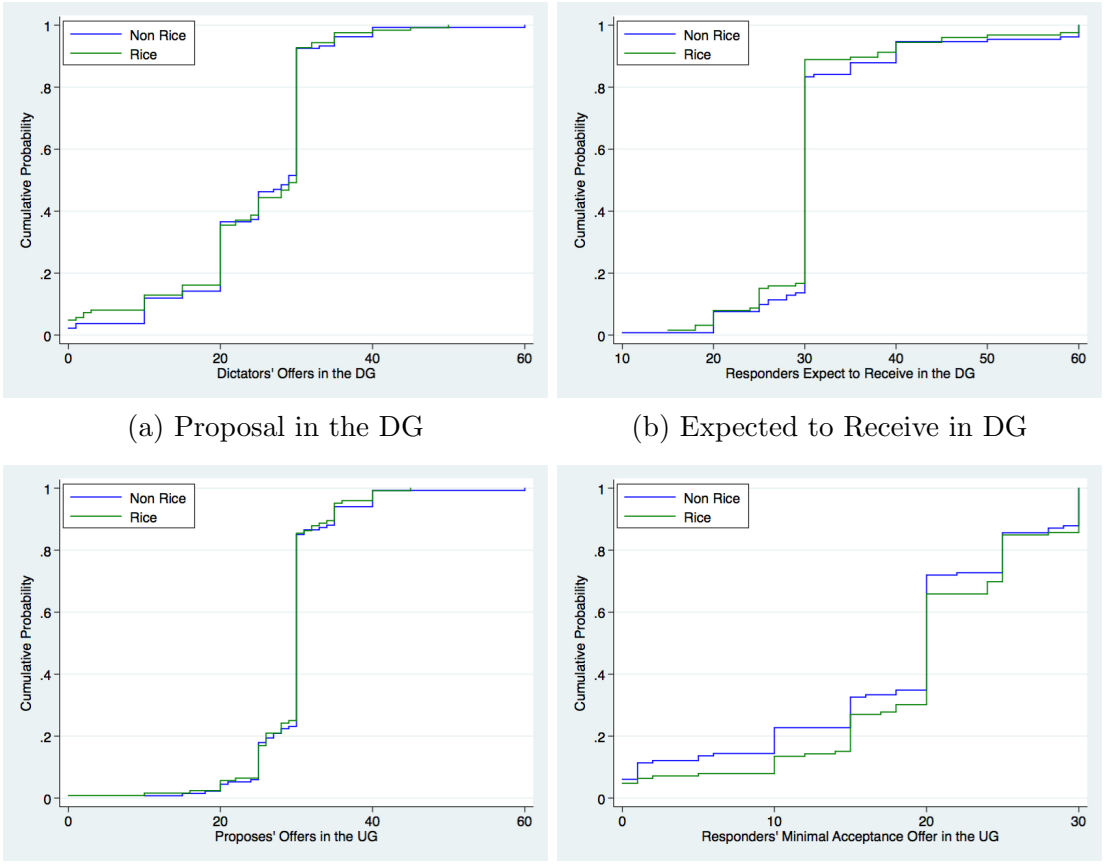
2.4.1 Rice Farming and Social Preferences

In the previous section, we have shown that rice subjects contribute more than non-rice subjects in the PGGs, and rice subjects are more likely to punish free-riders. There are a number of well known social preference models that can potentially rationalize these differences. These include: inequality aversion models (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), models that incorporate social welfare or efficiency concerns (Charness and Rabin, 2002), and reciprocity models (Rabin, 1993; Falk and Fischbacher, 2006). In this subsection, we discuss the explanatory power of each model.

Results from the dictator game and ultimatum game, which are presented in figure 2.3, suggest that pure altruism and inequality aversion is not the main difference between rice and non-rice subjects. Panel (a) and panel (c) of figure 2.3 present the distribution of offers in the dictator game and ultimatum game respectively. The distribution between rice and non-rice is very similar (Kolmogorov-Smirnov test: $p > 0.95$ for dictator game and ultimatum game). Responder's behavior in the two games is presented in panel (b) and (d). There is also no significant difference between rice and non-rice subjects (Kolmogorov-Smirnov test: $p > 0.66$ for the dictator game and ultimatum game).

In reciprocity models, beliefs about other's behavior play an important role. We did not elicit subjects' beliefs or ask subjects to make a contingent contribution plan based on other group member's contribution as in Fischbacher et al. (2001). This is because we employed a repeated PGG and each session already lasts about 2 hours. Nevertheless, in the following, we present four pieces of evidence showing that the difference in beliefs might not be an important factor in explaining the results.

Figure 2.3: Behavior in the Ultimatum Game and Dictator Game



(a) Proposal in the DG (b) Expected to Receive in DG

(c) Proposal in the UG (d) Responder Behavior in the UG

Notes: In the DG, while proposers were making decisions, responders were asked to specify the amount they expect to receive from the proposer. Responders knew that this would not affect the outcome of the dictator game and the proposers were not informed about this procedure. In the UG, responder's minimal acceptable offer was elicited. Before knowing the actual offers made by proposers, responders were asked to specify a number, which is the Minimal acceptable offer. If the proposer's offer is higher than or equal to this number, the offer is automatically accepted. Otherwise, the offer is automatically rejected.

First, if rice and non-rice subjects hold different beliefs, we should observe that rice's contribution is different to non-rice in the very first period of the PGG. We have already shown in the previous section that the difference is very weak. Moreover, the difference in the second period is also small (Mann whitney U test, each group as an independent observation: $p = 0.07$). Second, in the dictator game, the amount responders expect to receive from the dictators does not differ between rice and non-rice regions. Third, in the post experimental questionnaire, we asked subjects to what extend they believe that people from the local province obey public order, for example, do not jump queues, do not spit, and do not shout in public areas (10

points scale, 1 = completely disobey, 10 = completely obey). The average score in rice and non-rice regions is not significantly different from each other (Mann whitney U test: $p > 0.21$). Last but not least, the proportion of the efficient choice in the stag hunt game is very similar between rice and non-rice subjects (Two-sample test of proportions: none rice = 72.7%; rice = 72.4%; $p > 0.94$). Please note that 'hare' in the stage hunt game is a risk dominance action, which means if one is uncertain of other player's strategy, the expected payoff of selecting 'hare' is higher. On the other hand, 'stag' is the best response if one believes that the other player also choose the efficient action. Since both rice and non-rice subjects coordinate so well on the efficient outcome, it is reasonable to attribute this to the fact that they hold similar beliefs regarding the level of cooperativeness of other participants.

Reciprocity models also incorporate cases in which subjects reciprocate on final outcomes. Relating to the PGGs, subjects might contribute more if other group members contributed more in the previous periods. To test whether rice and non-rice subjects differ in this regard, we conduct two random effects panel regressions similar to Croson (2007).

The results in table 2.9 illustrate that there is no difference in the level of reciprocity between rice and non-rice subjects. Column 1 shows the results for the no punishment condition. The lag of other group member's Min, Median, and Max are all positively signifiant. However, regarding the interaction terms, only the "lag of others' max × Rice" is significant at 10% level. This suggests that rice subjects are slightly more responsive to the maximum of other group member's contribution. None of the interaction terms is significant in the punishment condition (column 2). This result alongside with the finding that rice and non-rice subjects hold similar beliefs suggest that the reciprocity models might not be able to account for the behavioral differences in the PGGs between the two regions.

2.4.2 Differences other than Rice Cultivation

The rice and non-rice provinces in our sample differ in other aspects besides the type of cultivation. Therefore, our finding might be a manifestation of the other differences instead of the difference in rice farming. In this subsection, we discuss several possible alternative explanations in detail and show that our results are less likely driven by other factors.

First of all, please note that a number of factors are already controlled by design. Nearly all of the subjects are first year, Han Chinese university students. Therefore, they all speak Mandarin, have similar education background, share the same culture origin (Wen et al., 2004), and live in the same political institution.[24]

Secondly, our results might be driven by the difference in economic development and market integration, since Henrich et al. (2010, 2001) find that these variables are positively associated with the level of cooperation and trust. In the previous section, we showed that our results are not affected by controlling for the municipality level GDP per capita as well as whether the subject come from rural or urban areas.

Third, climate might also have an effect on people's behavior. According to the pathogen prevalence theory, in regions that are prevalent of transmittable diseases, interaction with strangers are dangerous. Therefore, there is an evolutionary pressure for people living in those environments to form small and closed social groups, which consequently leads to the development of a more collective social norm (Fincher et al., 2008). Since rice areas in our sample are warmer than non-rice areas, and pathogens are highly correlated with temperature (Guernier et al., 2004), our results might be driven by the difference in climate instead of cultivation. Due to perfect collinearity, we are unable to control for temperature. However, in the following, we present two pieces of evidence to show that our results are less likely driven by climate difference. First, we administrated the Individualism and Collectivism questionnaire and find no difference between rice and non-rice subjects. We also controlled for subject's level of collectivism in the regression analyses, and the results are not affected. Second, all the four provinces have similar climate. According the Koppen climate classification, which is the mostly widely used method, three of the provinces in our sample belong to the same group C temperate climate category.

Lastly, if our results are driven by other cross province unobservable factors, we should observe differences between the two rice farming provinces as well as differences between the two non-rice farming provinces. The results in table 2.10 and table 2.11 suggest this is not the case. The results show that subjects from the two rice provinces behave similarly in the PGGs. This is also true for subjects from

[24]See Guiso et al. (2006) for a review of the effect of culture on economic outcomes. Chen (2013) find that the necessity to grammatically distinguish future and present events leads to more a present biased time preference. Bó et al. (2010) discover that people are more cooperative in social dilemma situations under democratic institutions.

the two non-rice provinces. In the appendix A.2, we also show that the probability of punishment as well as the intensity of punishment conditional on punishing are also similar among rice subjects (table A.4) and non-rice subjects (table A.5).

At the very least, we believe it is very unlikely that the factors that confound rice cultivation are only able to affect subject's behavior in the PGGs but not other measures, such as behavior in DG, UG, Stag Hunt, as well as the answers in the Triad task and Social styles questionnaires.

2.4.3 Self selection into rice and non-rice regions

Our main findings might also be driven by sorting. In particular, if a family does not like the social norm of a rice region, they can move to a non-rice province, or vice versa. We address this issue by further excluding observations whose father's birth place was a rice province but themselves have Hukou from a non-rice province or vise versa. Differently put, we drop subjects who had likely been migrated into or out of rice farming regions.

This strategy can control for the self selection issue because prior to 1990, the Chinese government had enforced laws that restricted immigration and travel within China.[25] Also note that China has strict immigration control since the Qin Dynasty (221 to 206 BC). This is because people are a scare resource. Both women and men are needed to tend the farming lands and men are required to form armies (See Jinguang (2004) for the documentation of the policy in Qin Dynasty; see the Book of Han for immigration policy in Han Dynasty; See the Tang Code for the policy in Tang Dynasty; See the Collected Regulations of the Great Ming for the policy in Ming Dynasty.)

We did not ask subjects for their father's age, but it is reasonable to assume that their father had born before the time when the immigration law was lifted.[26] Consequently, the birth place of subject's father was not due to selection, and subjects who still have the same Hukou as their father's should be free from selection issues. About 8 subjects are dropped, and the results are not affected by excluding them. As a further check, we drop 9 subjects whose *grow up* province has a different

[25] See Qian (2008) and the references therein for the details about the policy.

[26] The normal age for the first year undergraduate student is 18 years old. Therefore, their parents had born before the lift of the immigration law as long as they had their children after 8 years old.

type of cultivation than their father's birth place. Our results are also robust in this specification.

2.5 Discussion

In the previous sections, we show that rice subjects contribute more than non-rice subjects in the PGGs with and without punishment, with the difference much stronger in the punishment condition. We also find that rice subjects are more likely to punish free-riders, and conditional on punishing, there is no difference in the intensity of punishment. Based on these results, we argue in the following that the practice of rice farming probably does not change people's preference for cooperation. Instead, is it more likely the case that people in rice areas understand the public goods situation differently than their non-rice counterparts.

First, if individuals from rice farming regions have a preference for cooperation, they should contribute more than non-rice subjects in the no punishment condition. However, this is not the case. The difference over the eight periods is relatively weak, and the contribution level is the same in the first period.

Second, the possibility of future interaction is crucial for cooperation to emerge in the absent of punishment opportunities (Rand and Nowak, 2013), and rice subjects seem to understand this feature. According to the contribution pattern in the no punishment condition (period 1-8 of figure 2.1), rice subject's contribution increases at the early periods when there is more room for future interaction. However, their cooperation level tumbles down in later periods. On the other hand, the contribution pattern of non-rice subjects are relatively flat over interaction (see table 2.2).

Third, rice subjects are more willing to punishment free-riders, which is one of the most effective way to foster cooperation (Fehr and Gächter, 2000; Masclet et al., 2003). Moreover, there is no difference between rice and non-rice subjects in punishing cooperators, which is detrimental for cooperation (Herrmann et al., 2008). We also find that the difference in pro-social punishment mainly comes from the fact that rice subjects are more likely to punish free-riders, and conditional on punishing, there is no difference in the intensity of punishment. Rice subject's punishing behavior is efficient because more punishment points assigned to free-riders does not lead to higher cooperation level (Result 5 shows that rice and non-rice subjects respond similarly to punishment), the act of punishment alone is sufficient

to boost cooperation.

The reason that people from rice farming areas are more willing to exercise their right to punish is probably the fact that punishment is easy to enforce in rice farming environments. First, since rice farming requires a large amount of labour, each farmer benefits from successful cooperation. This feature ensures that free-riding is socially undesirable in the rice farming societies. Second, the paddy fields are usually plain lands without any shelter, therefore, each farmer's effort is easily observable. Third, the cost of avoiding punishment or choosing the "outside option" is extremely high for farmers, since they will loose their farming land if they migrate to another area. Last but not least, the names of the free-riders might spread relatively fast within the village, since rural villages are usually small and closed communities. These characteristics are in line with Debraj's summarization of the broad conditions that need to be satisfied for punishment behavior to occur. The conditions include: positive individual gain from successful cooperation, member's action must be observable by others, and sanctions must be enforceable ((Ray, 1998)).

2.6 Conclusion

This paper explores the origins of the cross cultural differences in the level of cooperativeness. In particular, we show that rice cultivation that practiced hundreds of years ago leads to a cooperative social norm that is intergenerational transmittable. To test our hypothesis, we travelled to four typical rice and non-rice farming provinces in China and recruited local *university students* who had minimal farming experience as subjects. We find that compared to subjects from non-rice regions, subjects from rice areas contribute more in both the no-punishment and punishment conditions of the public goods game. Rice subjects are also more likely to punish free-riders and conditional on punishing, there is no difference in the amount of punishment. Importantly, we did not find any difference between rice and non-rice subjects in games that do no involve cooperation, such as the DG, UG, and Stag Hunt game. These results suggest that rice cultivation might not change individual's preference for cooperation. Instead, is it more likely the case that people in rice areas understand the public goods situation differently than their non-rice counterparts.

Table 2.8: Regressions regarding the Probability of Punishment and the Intensity of Punishment Conditional on Punishing using Municipality level Rice statistics

	Probability of Punishment			Intensity of Punishment		
	(1)	(2)	(3)	(4)	(5)	(6)
Perc. Paddy Field (Municipality)	0.000945*	0.00120**	0.00111**	0.00610*	0.00585*	0.00433
	(0.000508)	(0.000496)	(0.000478)	(0.00340)	(0.00325)	(0.00306)
Anti-Social Punishment	-0.137***	-0.0813***	-0.0809***	-0.106	0.358	0.279
	(0.0156)	(0.0193)	(0.0192)	(0.166)	(0.248)	(0.249)
Perc. Paddy Field X Anti-Social Punishment	-0.000206	-0.000105	-0.000105	-0.000805	-0.000186	0.00217
	(0.000321)	(0.000319)	(0.000318)	(0.00405)	(0.00415)	(0.00394)
Punisher Contribution		0.0107***	0.0106***		0.0598	0.0396
		(0.00306)	(0.00308)		(0.0368)	(0.0312)
Punished Contribution		-0.0102***	-0.0102***		-0.0373*	-0.0397**
		(0.00199)	(0.00199)		(0.0192)	(0.0171)
Other Two Member Avg Contribution		-0.0187***	-0.0186***		-0.0249	-0.0310
		(0.00307)	(0.00308)		(0.0293)	(0.0258)
GDP per cap. (Municipality)			-0.000158			0.00298
			(0.00456)			(0.0278)
Holistic Thinking			-0.145**			0.314
			(0.0660)			(0.417)
From Rural (Dummy)			0.000113			0.296
			(0.0337)			(0.197)
Relative Income			0.00170			0.133
			(0.0209)			(0.142)
Risk Attitude			0.0102			0.140*
			(0.00787)			(0.0778)
Male			0.0191			0.690***
			(0.0283)			(0.247)
Natural Science (Dummy)			-0.00242			0.219
			(0.0259)			(0.169)
Priming (Dummy)			0.0825**			-0.119
			(0.0357)			(0.215)
Single Child (Dummy)			-0.0158			-0.139
			(0.0318)			(0.245)
Collectivistic			0.000961			0.0705
			(0.0338)			(0.253)
Individualistic			0.0634*			0.398*
			(0.0349)			(0.216)
Trustworty (Belief)			-0.00880			0.0246
			(0.00810)			(0.0589)
Public Order (Belef)			0.00125			0.0175
			(0.00609)			(0.0511)
Period	-0.0232	0.0198	0.0206	0.108	0.153	0.0271
	(0.0458)	(0.0454)	(0.0453)	(0.291)	(0.286)	(0.283)
Period Squared	0.000549	-0.000961	-0.000986	-0.00622	-0.00781	-0.00236
	(0.00182)	(0.00180)	(0.00179)	(0.0117)	(0.0115)	(0.0112)
Constant	0.549**	0.440	0.312	1.680	1.158	-1.183
	(0.278)	(0.277)	(0.307)	(1.803)	(1.802)	(2.169)
sigma						
Constant				1.864***	1.852***	1.754***
				(0.168)	(0.169)	(0.137)
-						
Perc. Paddy Field (Municipality)						
Observations	7519	7519	7519	2379	2379	2379

Notes: Columns 1-3 are Random Effects Linear Probability models and the dependent variable is the probability of punishing Free-rider and Cooperators. Columns 4-6 are tobit regressions and the dependent variable is the amount of punishment points assigned to other group member who contributed more or less than the punisher. Perc. Paddy Field is the percentage of cultivated land devoted to paddy fields at the municipality level. Cluster Standard errors (PGG group level) are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

Table 2.9: Random Effects Panel Regressions about reciprocity in the PGGs

	(1) No Punishment	(2) Punishment
Rice	0.0991	0.432
	(0.649)	(0.775)
Other Median (Lag)	0.152***	0.442***
	(0.0503)	(0.0606)
Rice x Other Median	-0.0856	-0.0139
	(0.0715)	(0.0714)
Other Max (Lag)	0.154***	0.130**
	(0.0288)	(0.0510)
Rice x Other Max	0.0824*	0.0480
	(0.0481)	(0.0710)
Other Min (Lag)	0.251***	0.141***
	(0.0465)	(0.0376)
Rice x Other Min	0.0329	-0.0373
	(0.0631)	(0.0483)
Period	0.675***	-0.276
	(0.201)	(0.317)
Period square	-0.0944***	0.00818
	(0.0205)	(0.0120)
Punish Receive (Lag)		-0.290***
		(0.0638)
Constant	3.827***	6.068***
	(0.563)	(2.148)
Observations	3619	3619

Notes: The dependent variable is contribution in the PGGs. Other Max (lag) is the maximum of other group members contribution in the previous period. Other Median and Min (Lag) are defined analogously. Cluster Standard errors (PGG group level) are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

Table 2.10: Random Effects Panel Regressions Comparing contribution in the PGG with punishment between the two rice provinces and the two non-rice provinces.

	Non-Rice Provinces		Rice Provinces	
	(1)	(2)	(3)	(4)
Shandong	-1.114	-0.534		
	(0.890)	(0.781)		
Zhejiang			0.531	0.319
			(0.902)	(0.989)
GDP per cap. (Municipality)		0.115		-0.0543
		(0.0736)		(0.0911)
Holistic Thinking		0.283		0.373
		(0.804)		(1.164)
From Rural (Dummy)		-0.278		-0.609
		(0.710)		(0.526)
Relative Income		-0.533*		0.187
		(0.279)		(0.364)
Risk Attitude		0.0525		0.261*
		(0.159)		(0.148)
Male		2.142***		1.947***
		(0.575)		(0.460)
Natural Science (Dummy)		1.771***		0.604
		(0.513)		(0.557)
Priming (Dummy)		-0.533		0.286
		(0.738)		(0.858)
Single Child (Dummy)		-0.172		0.311
		(0.623)		(0.570)
Collectivistic		0.941		0.187
		(0.625)		(0.673)
Individualistic		-0.0409		-0.611
		(0.557)		(0.546)
Trustworty (Belief)		0.208		-0.00377
		(0.133)		(0.148)
Public Order (Belef)		-0.143		0.0928
		(0.105)		(0.117)
period		1.875***		2.633***
		(0.386)		(0.468)
period × period		-0.0623***		-0.0876***
		(0.0144)		(0.0183)
Constant	12.49***	-6.757	13.16***	-6.853*
	(0.729)	(4.661)	(0.645)	(3.929)
Observations	2112	2112	2000	2000

Notes: the dependent variable is the contribution in the PGG with punishment. The results are the same for the no punishment condition. We did not include it because the difference is weak between rice and non-rice. Cluster Standard errors (PGG group level) are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

Table 2.11: Tobit Regressions Comparing punishment behavior between the two rice provinces and the two non-rice provinces.

	Non-Rice Provinces		Rice Provinces	
	(1)	(2)	(3)	(4)
Shandong	-0.163	0.278		
	(0.408)	(0.382)		
Zhejiang			0.167	-0.146
			(0.634)	(0.731)
Anti-Social Punishment	-1.471***	-1.312***	-1.918***	-1.743***
	(0.315)	(0.257)	(0.458)	(0.450)
Shandong x Anti-Social Punishment	0.246	0.0129		
	(0.457)	(0.402)		
Zhejiang x Anti-Social Punishment			0.0114	0.0829
			(0.601)	(0.549)
GDP per cap. (Municipality)		-0.0902		0.0108
		(0.0565)		(0.0686)
Holistic Thinking		-1.430**		-1.516*
		(0.646)		(0.853)
From Rural (Dummy)		0.107		0.218
		(0.362)		(0.392)
Relative Income		-0.0714		0.357
		(0.298)		(0.264)
Risk Attitude		0.227*		-0.0524
		(0.123)		(0.143)
Male		-0.128		1.040**
		(0.275)		(0.507)
Natural Science (Dummy)		-0.173		0.0946
		(0.351)		(0.380)
Priming (Dummy)		0.923**		0.155
		(0.410)		(0.588)
Single Child (Dummy)		0.0293		-0.563
		(0.382)		(0.515)
Collectivistic		-0.0617		0.360
		(0.517)		(0.489)
Individualistic		0.811*		0.150
		(0.429)		(0.520)
Trustworty (Belief)		0.0252		-0.130
		(0.102)		(0.113)
Public Order (Belef)		-0.0802		0.0761
		(0.0773)		(0.0979)
period	0.157	0.163	-0.741	-0.757
	(0.542)	(0.510)	(0.722)	(0.707)
period × period	-0.0132	-0.0132	0.0255	0.0264
	(0.0222)	(0.0209)	(0.0284)	(0.0278)
Constant	-0.901	-2.811	4.339	3.376
	(3.179)	(3.469)	(4.484)	(4.890)
Observations	3985	3985	3534	3534

Notes: rhe dependent variable is the amount of punishment points assigned to free-riders or cooperators. Cluster Standard errors (PGG group level) are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

Chapter 3

Does testosterone foster reputation concerns?

3.1 Introduction

Testosterone is a steroid hormone that plays a critical role in brain development and sexual behavior for both males and females. Early studies, using animals as subjects, have established a causal link between testosterone and aggressive behavior (Edwards, 1969; Beeman, 1947). However, recent evidence suggests that the role of testosterone on human's behavior is more complicated. It causes aggressive behavior only when individual's status is being challenged (Eisenegger et al., 2011; Mazur and Booth, 1998). These findings lead researchers to hypothesize that testosterone's main role in human social interaction is to promote dominance behavior, that is, actions to achieve or maintain high social status (Mazur and Booth, 1998; Josephs et al., 2006; Terburg et al., 2009; Josephs et al., 2003).

There is mounting evidence in the psychology and neuroscience literature supporting the dominance hypothesis. Testosterone has been shown to reduce unconscious fear (van Honk et al., 2005), foster responsiveness to angry faces (van Honk et al., 2001), and increase vigilance to social threats (van Honk et al., 1999). These findings suggest that one important role of testosterone is modifying human's physical and psychological fitness so that they are more prepared for impending social challenges. Additionally, van Honk et al. (2011) find that testosterone impairs cognitive empathy. Empathy means that people are concerned with the emotions or

welfares of others. However, this concern might cause adverse effects when competing with rivals (Eisenegger et al., 2011). Last but not least, results from Bos et al. (2010) and Boksem et al. (2013) suggest that testosterone decreases trust. This finding is also consistent with the dominance hypothesis, because the decision to trust others bears the risk that the trust been exploited and thus give the betrayer a relative advantage in the competition for status.

Although the aforementioned evidence does not link testosterone to aggressive behavior per se, it increases the chance of social conflict. Note that social status can also be obtained by prosocial behavior in situations where there is no potential social challenges (Anderson and Kilduff, 2009). However, to the best of our knowledge, no study has investigated whether testosterone leads to status enhancing behavior even in the absence of status threats.

In this paper, we fill this gap by testing whether testosterone leads to prosocial behavior in the context of charitable donations. We believe charitable donation is a perfect environment for the hypothesis because numerous studies have demonstrated that reputation or status concern is an important motivation underlying individuals' giving behavior (see Harbaugh (1998); Glazer and Konrad (1996) for theoretical models and Andreoni and Petrie (2004); Rege and Telle (2004); Van Vugt and Hardy (2010); Ariely et al. (2009); Kataria and Regner (2015); Karlan and McConnell (2014); Soetevent (2005); Alpízar et al. (2008); Alpízar and Martinsson (2013); List et al. (2004) for empirical evidence.)

Seventy-seven healthy white female participated in our study. A single dose of 0.5 mg of testosterone or placebo was applied sublingually in a randomized and double-blind fashion. Only females were recruited because it takes about 4 hours for a single sublingual administration of 0.5mg of testosterone to have a neurophysiological effect. The time is unknown for males (Tuiten et al., 2000). Subjects were asked to work on a real effort task to generate money for their chosen charity. They were not allowed to earn money for themselves in the real effort task because first, monetary incentives might dilute the status value of donation (Ariely et al., 2009; Bénabou and Tirole, 2006), and second, financial payments might add another layer of complexity to subject's behavior, because studies have discovered that testosterone makes people more motivated towards monetary rewards (Eisenegger et al., 2011; Hermans et al., 2010).

Subjects' reputation concerns were manipulated in three experimental condi-

tions. In the honor condition, subjects who donated more than an exogenously set threshold were publicly acknowledged. In the shame condition, the subject who donated the least was identified in public.[1] Subject's effort in the two conditions was compared to the baseline condition, in which all the donations were kept in private. Our experimental setting combines both the within and between subjects design features. The between element is the random administration of testosterone and placebo; the within feature is that subjects participated in both baseline and honor conditions (refer as honor treatment) or baseline and shame conditions (refer as shame treatment) in each session. The honor treatment is a direct test of our hypothesis because subjects can gain reputation by achieving a certain amount of donation, and more importantly, there is no risk of losing of reputation. On the other hand, the situation in the shame treatment is more related to the dominance hypothesis since being publicized as the least donor is a loss of status.

We find that there is no behavioral differences between the placebo and testosterone group in both the honor and shame treatments. In the discussion section, we argue that three factors might contribute to the null finding. First, the real effort task might be too enjoyable. Second, the subjects were all female subjects. And last, the null result might due to lack of statistical power. Power analyses conducted in the experimental design section suggest that the shame treatment has enough statistical power to detect a large difference, but is underpowered for small effects; the honor treatment, on the other hand, is underpowered even for large differences.

To the best of our knowledge, we are the first to investigate whether testosterone induces reputation seeking behavior in the absence of social threats. The two studies that are closely related to ours are Eisenegger et al. (2010) and van Honk et al. (2012). Eisenegger et al. (2010) find that testosterone causes subjects to make fairer offers in the ultimatum game. Note that in bargaining situations, individual's social status can still be challenged. This is because the proposer faces the risk that her offer might be rejected by the responder, which is considered as a social affront. A higher or fairer offer significantly reduces the probability of rejection. Moreover, one possible alternative explanation to their finding is that testosterone might make people care about money, and hence subjects offer a fairer amount to the responder to reduce the risk of rejection, in which case both the proposer and responder earn

[1]The reason that we switched from a absolute threshold in the honor condition to a relative threshold in the shame condition is discussed in the experimental design section.

nothing.

van Honk et al. (2012) find that subjects who intake testosterone contribute more than the placebo group in a public goods game. However, it is hard to attribute their findings to the effect of testosterone on reputation concerns. The reason is that all the decisions were made in strict anonymity. The present study also contributes to the literature that investigates the role of testosterone on human social interactions (for a brilliant review see Eisenegger et al. (2011)).

The rest of the paper is organized as follows. Section 3.2 explains the experimental design. The hypotheses are proposed in section 3.3. Section 3.4 illustrates the results. Section 3.5 discusses several possible explanations of why we find a null result, and Section 3.6 concludes.

3.2 Experimental Design

Seventy-seven healthy white female participants with an age range of 18 to 35 (mean = 23.7) participated in the study. They were recruited via on campus advertisement and mailing list at the University of Cape Town and were paid a flat fee of 200 RANDS. They were included in the study after signing a written consent and none of the them withdrew in the middle of the experiment. The study was approved by University of Cape Town Research Ethnics Committee (HREC REF 092/2011).[2]

A single dose of 0.5 mg of testosterone or placebo was applied sublingually in a randomized and double-blind fashion. Only females were recruited because it takes about 4 hours for a single sublingual administration of 0.5mg of testosterone to have a neurophysiological effect, and the time is unknown for males (Tuiten et al., 2000).

Power analyses suggest that we need a total of 52 subjects to have a power of 0.8 to detect a large effect under the 5% significance level. Since there were seventy-seven subjects participated in the study, we have enough observation if we combine both of the experimental treatments. The treatments are honor treatment and shame treatment, in which we investigate the effect of testosterone on subject's motivation on the pursuit of honor and the avoidance of shame, respectively.[3] We

[2]We were careful in our wording for the shaming procedure in order not to make them uncomfortable. Note that we are interested in how subject respond to the fact that the least donator will be shamed, not the shaming itself. Therefore, the shaming is quite weak at the end of each session.

[3]In this case, the hypothesis is that whether testosterone affect's people's reputation concerns

have a total of 48 subjects in the shame treatment, which means that we have the statistical power to detect a large difference for the shame treatment. However, the honor treatment is underpowered to detect a large difference as there were 29 participants.

3.2.1 The real effort task

In the experiment, subjects worked on a real effort task in order to generate money for charity organizations. We implemented a donation environment because numerous studies, both theoretical and empirical, have shown that social status is an important motivation underlying people's charitable behavior (see Harbaugh (1998); Glazer and Konrad (1996) for theoretical models and Andreoni and Petrie (2004); Rege and Telle (2004); Van Vugt and Hardy (2010); Ariely et al. (2009); Kataria and Regner (2015); Karlan and McConnell (2014); Soetevent (2005); Alpízar et al. (2008); Alpízar and Martinsson (2013); List et al. (2004) for empirical evidence). We employed a real effort task instead of a monetary windfall because firstly, we believe that real effort task is more similar to real life situations, in which individuals need to work to earn money and donate part of their earnings to charity. Secondly, studies have documented that subjects behave more selfishly when their endowment was earned via a real effort task comparative to windfall money in dictator games (Cherry et al., 2002) and in the context of charitable donations (Carlsson et al., 2013). Thus, we believe using a real effort task offers a more stringent test of our hypotheses.

The real effort task is called find the 'T's. When the task starts, a page full of symbols is presented on subject's computer screen (figure 3.1). There are only two different symbols, one is '+' (panel A of figure 3.2) and the other is more similar to the letter 'T' (panel B of figure 3.2). There are 20 'T's and 60 '+'s on each page. Participants are given 15 minutes to find as many 'T's as possible. Once all the 20 'T's have been found, a new page of symbols will appear. Subjects are informed that they do not need to perform the task for the full fifteen minutes; they can stop at any time they want. We also provide subjects with the option that they can

without further distinguishing between the pursuit for honor and the avoidance of shame. Results suggest that the difference is not significant even if we combine the two treatments (Mann-Whitney U test $p > 0.89$).

continue to work on the task but no money will be generated.[4] In order to prevent random clicking without distinguishing '+' and 'T', we impose the following rule: Subjects are allowed to make no more than 3 mistakes on each page. If they make the forth mistake all 'T's found on that page will be deselected, which means they need to restart the page again. However, this does not affect previously solved pages. Additionally, the number of mistakes on a page do not carry over to the next page. In other words, once they have completed a page, the number of mistakes will be reseted to zero.

We employed a diminishing payment scheme: 162.5 cents are donated for each of the first 20 'T's, 78 cents each for the next 20 'T's, 51 cents each for the next 20 'T's, ..., and 3.15 cents for each 'T' above 380.[5]

Importantly, subjects were not allowed to earn any money for themselves in the real effort task. In other words, all the money generated from the real effort task was donated to charity. The reasons for this design are the following. First, the effect of testosterone on reputation concerns might be weakened by the desire to earn money, because studies have shown that testosterone makes people more motivated towards monetary rewards (Eisenegger et al., 2011; Hermans et al., 2010). Second, monetary incentives might dilute the status value of donation, as observers cannot distinguish whether the higher donation is due to good deeds or simple to earn more money (Bénabou and Tirole, 2006; Ariely et al., 2009). [6]

3.2.2 Experimental Conditions and Treatments

Subjects' image concerns were manipulated in three experimental conditions. In the honor condition, subjects who donated more than an exogenously set threshold were publicly acknowledged.[7] In the shame condition, the subject who donated the

[4]We provide this option for subjects who are not interested in making donations but might be interested in playing the real effort task. Note that no subject chose this option.

[5]We use this diminishing payment scheme to increase the relative cost of making donations. Subjects who are less motivated should stop donating when the payoff is too low. Ariely et al. (2009) also used a diminishing payment scheme.

[6]We believe that investigating the effect of testosterone on individual's tradeoff between monetary payment and reputation concern is very interesting. Possible designs to better understand individual's behavior include treatments in which subjects can only earn money for themselves and can earn money for themselves as well as for charity. We leave these to future research.

[7]Note that in the honor condition, subjects knew that the honoring threshold is solving 26 pages. There is no such threshold in the baseline condition. In this case, there are two differences between the baseline and honor condition. The first is the effect of honor, which we want to

Figure 3.1: The real effort task – Find the 'T's

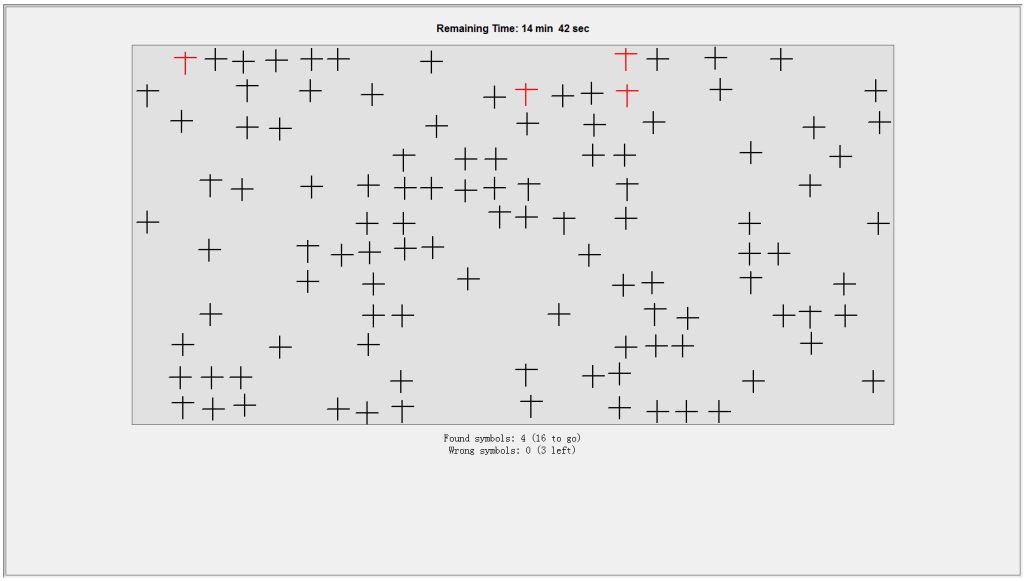
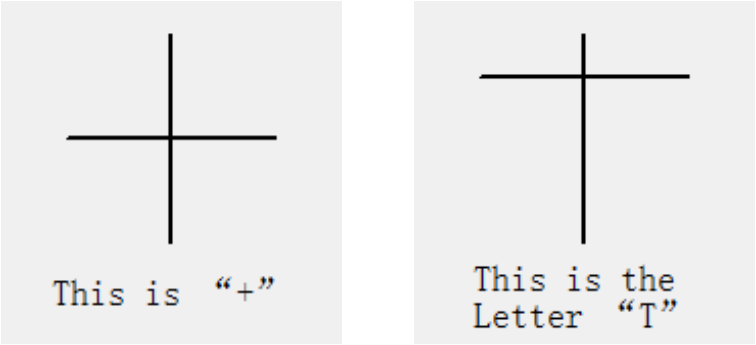


Figure 3.2: The real effort task – The different symbols



(a) (b)

least was identified in public. Subject's effort in the two conditions is compared to the baseline condition, in which all the donations were kept in private. Note that we switched from a absolute threshold in the honor treatment to a relative threshold in the shame treatment. The main reason for this modification is that we can set a relatively high threshold in the honor treatment because our main interest is how subjects respond to the image incentive. Whether or not they can achieve

identify. The other is the focal point, 26, which is only shown in the honor condition. In order to make the baseline and honor condition comparable, we also primed the threshold in the baseline condition. We stated in the instruction: "Note, earlier studies have shown that people can solve 19 tasks with little effort, and can solve 26 or more tasks with reasonable effort."

the threshold is not important. However, it is hard find a proper threshold for the shame treatment. If the threshold is too low, subjects can easily avoid shame. On the other hand, if the threshold is too high, subjects will know that all of them are going to be 'ashamed', which is not a 'punishment' at all.

Subjects in each session participated in both the baseline and honor conditions (the honor treatment) or the baseline and shame conditions (the shame treatment). Note that the order is always baseline – honor (shame).[8] In sum, we implemented a 2x2 between-within subjects design, with one treatment dimension as testosterone and placebo, and the other as honor and shame (see table 3.1)

Table 3.1: Summary of Experimental Treatments

	Placebo	Testosterone
Honor Treatment	Baseline – Honor	Baseline – Honor
Shame Treatment	Baseline – Shame	Baseline – Shame

The honoring and shaming procedure was very simple. Those who donated more than the threshold or donated the least in each session were asked to stand up. In the honor treatment, the experimenter announced: "These people have solved 26 tasks or more and hence donated the most, we would like to thank you for your efforts and we will make sure that your chosen charities will benefit from your effort." The line was "This person solved the least, nevertheless we would like to thank everyone, for your efforts" in the shame treatment.

3.2.3 Experiment Procedure

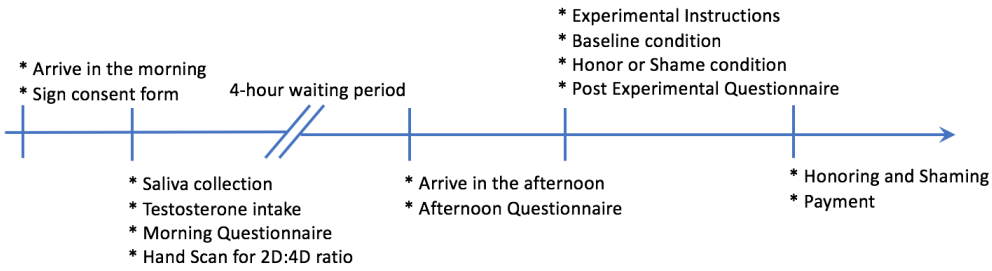
Participants arrived in the experimental lab between 9:00 and 12:00 in the morning. They provided their saliva sample after signing a written consent. Afterwards, they were instructed to intake either a testosterone or placebo pill. Participants then completed a questionnaire, and their hands were scanned for the 2D:4D ratio. They were free to leave and were required to return to the lab in the afternoon, about 4 hours after the testosterone intake.

In the afternoon, participants were asked to complete another set of questionnaires. The instructions for the first part of each treatment (the baseline condition)

[8]The small number of participants did not allow us to counter balance the order. Please see the hypotheses section for a discussion of how this will affect our results.

were subsequently distributed. Participants received the instructions of the second part of the treatment (honor or shame condition) after the conclusion of the first part. Subjects knew that there were more than one part in each session, but they were not informed about the content of each condition before the corresponding instructions were handed out. The honoring and shaming procedure was at the very end of the experiment, after they completed a post experimental questionnaire. See figure 3.3 for the sequence of the events in each session.

Figure 3.3: The Sequence of Events in Each Session



We conducted 6 sessions of Honor treatment between May and August 2013 and 10 sessions of Shame treatment between November 2013 and October 2014. All the sessions were conducted in the University of Cape Town, South Africa.[9]

3.3 Hypotheses

In the honor treatment, subjects can gain reputation by being publicly acknowledged as a generous donator. More importantly, there is no risk of losing status because no one will be 'ashamed'. Therefore, if testosterone makes people care about their reputation even in absence of potential social threats, the testosterone group should try harder to get acclaimed in the honor treatment. On the other hand, being publicized as the least donator is a loss of reputation. Therefore, according to the dominance hypothesis, we expect that the subjects who intake testosterone put more effort in the shame condition relative to the baseline condition to avoid social stigma.

[9]See Appendix B for the experimental instructions for the honor treatment. The instructions for the Shame treatment is very similar.

Subjects' reactions to the reputation incentives are captured by the treatment effect, defined as the number of 'T's find in the honor or shame condition minus the number of 'T's find in the baseline condition. We use the treatment effect as our main dependent variable because subjects participate in both baseline and honor (shame) conditions in each session (the with-in element of our experiment). Thus, it has the advantage that some individual specific unobservables, such as sympathy towards their chosen charity, can be cancelled out.

Since reputation concern is an important motivation underlying charitable behavior, we expect the treatment effect in the honor and shame treatment is positive for both the testosterone and placebo group. Our main hypothesis that the treatment effect is larger for the testosterone group, since studies have shown that testosterone is a social hormone and makes people care more about their social status. One caveat is that the order of the conditions is always baseline – honor (or shame). Consequently, the higher donation in the honor and shame conditions relative to the baseline condition might simply due to learning. We want to point out, however, that the effect of learning presents in both the testosterone and placebo group. Hence, the difference in treatment effect between the two groups is still able to identify the effect of testosterone on individual's reputation concerns.[10]

Hypothesis 1. *The treatment effect is significantly higher for the testosterone group compared to the placebo group for both honor and shame treatments.*

Studies have discovered that the neurophysiological effect of exogenously administration of testosterone is positively correlated with testosterone exposure during the prenatal stage (van Honk et al., 2012, 2011). The intensity of prenatal testosterone exposure is marked by the ratio of the length of the index finger to the length of the ring finger (2D:4D) (Lutchmaya et al., 2004; Manning, 2002). In particular, the lower the 2D:4D ratio (index finger shorter than the ring finger) the stronger the exposure to testosterone in the uterus. Relating these findings to our experimental setting, subjects with a lower 2D:4D ratio are more influenced by testosterone and hence have a larger treatment effect.

Hypothesis 2. *The effect of testosterone intake on the treatment effect is stronger*

[10]Our results might be biased if testosterone affect's subject's learning rate of the real effort task. However, to the best of our knowledge, we do not find any evidence supporting the idea that testosterone fosters or hinders learning in real effort tasks that are similar to the task implemented in the present study.

for subjects with low 2D:4D ratio (high prenatal exposure to testosterone) compared to subjects with high 2D:4D ratio (low prenatal exposure to testosterone) for both honor and shame treatments.

2D:4D also has a non-monotonic influence on altruism. Brañas-Garza et al. (2013) find that subjects with both high and low 2D:4D are less altruistic than subjects with intermediate values of the ratio. Relating this finding to the present study, testosterone should have less effect on subjects with intermediate 2D:4D ratio (altruistic individuals) because sympathy towards the recipients will motivate them to donate a higher amount even in the absense of testosterone.[11]

Hypothesis 3. *The effect of testosterone intake on the treatment effect is stronger for subjects who have high or low 2D:4D (less altruistic) compared to subjects with intermediate 2D:4D values (more altruistic).*

The effect of testosterone might also interact with individual's inherent sensitivity to shame. In particular, in the shame treatment, testosterone might be less effective on subjects who dislike being publicly 'ashamed'. Subject's level of shame aversion is measured by the Social Desirability scale (Paulhus, 1984).[12] The questionnaire assesses people's tendency to make socially desirable responses to questionnaire questions, which also reflects people's sensitivity to shame. This is because if one is not motivated by image concerns, there is no need to report socially acceptable answers to sensitive questions such as drug use or domestic violent behavior.

Hypothesis 4. *In the shame treatment, subjects who have a inherently high level of shame aversion (score higher in the Social Desirability scale) should be less affected by testosterone in comparison to subjects who are less sensitive to shame.*

In the field of psychology, the canonical model of human behavior suggests that there are two motivational systems underlying individual's action (Carver and White, 1994; Elliot, 2006; Gray, 1990). One is the behavioral activation system (BAS), which motivates people when there are cues for reward. The other is the behavioral inhibition system, which influences behavior in situations with signs of

[11] We also administrated the altruism scale from the International Personality Item Pool. The results are similar.

[12] The questionnaire was taken in the morning part of the experiment, before the testosterone intake.

impeding punishment. Jeong et al. (2011) find that BIS/BAS is correlated with people's giving behavior. In particular, they show that subjects who are more sensitive to rewards (high BAS) donate a larger amount if the plead is framed positively, i.e,. "...With funds, it will be able to stay open longer hours for student use and expand the book collection." Similarly, subjects who are sensitive to signs of potential punishment (high BIS) donate more under the negatively framed message, i.e., "...Without funds, it will have to cut down on menu items and increase food prices."

In our experimental setting, the honoring is essentially a potential reward, therefore, subjects with high BAS should be less affected by testosterone in the honor treatment compared to subjects who are less sensitive to rewards (low BAS). This is because subjects who are more reward oriented are already highly responsive to the honor incentive even without the testosterone intake. The finding from Hermans et al. (2010) is also inline with the hypothesis. The authors find that the effect of testosterone on reward responsiveness is stronger for subjects with low BAS. On the other hand, shaming is a potential punishment. Similar arguments as in BAS lead to the hypothesis that subjects with high BIS are less affected by testosterone in the shame treatment compared to subjects who are less sensitive to punishments (low BIS). The BIS/BAS scale is from Carver and White (1994) and the items are coded so that higher BIS and BAS indicates higher sensitivity to punishment and reward stimuli.[13]

Hypothesis 5a. *The effect of testosterone administration on the treatment effect is lower for subjects who are more motived by BIS in the shame treatment.*

Hypothesis 5b. *The effect of testosterone administration on the treatment effect is lower for subjects who are more motived by BAS in the honor treatment.*

3.4 Results

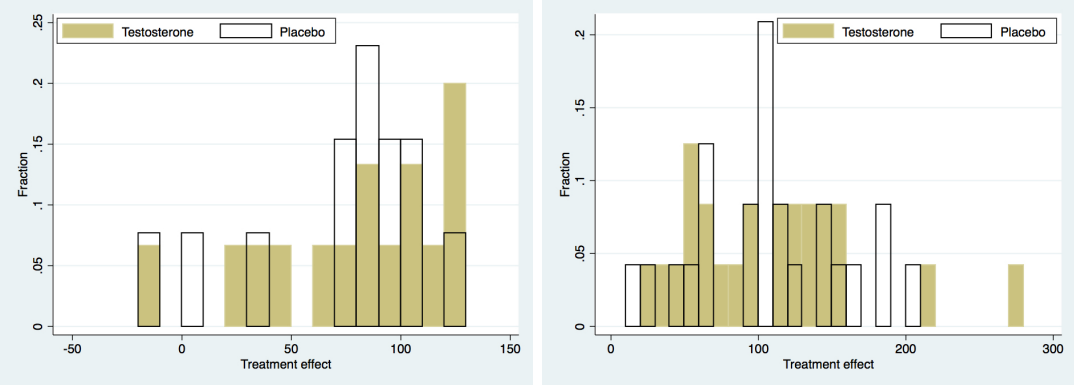
Figure 3.4 shows the histogram of the treatment effect, defined as the number of 'T's find in the honor or shame condition minus the number of 'T's find in the

[13]We use drive and reward seeking of the BAS scale in the analyses. Results are similar if use the all the sub-scales of BAS.

baseline condition.[14] The distribution of treatment effect is very similar between the testosterone and placebo group for both the honor (a) and shame (b) treatments (Mann-Whitney U tests, $p > 0.61$ for both shame and honor treatments). These findings are in contrast with the first hypothesis, which states that treatment effect should be higher for the testosterone group.

Note that subject's effort in the baseline treatment is similar between the Testosterone and Placebo group (Mann-Whitney U test, $p > 0.84$. We pooled the baseline treatment from the honor and shame treatment because subjects were unaware of the honor or shame condition when performing the baseline condition.) Since there is no reputation incentive in the baseline treatment, and subjects are not informed about the content of the next part of the experiment, this finding provides suggestive evidence that testosterone does not make individuals less altruistic.[15]

Figure 3.4: Histogram of the treatment effect in honor and shame treatments



(a) Honor treatment (b) Shame treatment

Notes: Treatment effect is defined as the number of Ts found in the honor or shame condition minus the number of Ts found in the baseline condition.

The results regarding hypothesis 2 are presented in figure 3.5. The figure shows the effect of testosterone on the treatment effect, broken down by a median split of the 2D:4D ratio (we use the median split to reduce noise of the data). As discussed in

[14] We dropped a subject who made an extremely negative treatment effect in the honor treatment. She stated that she felt very nervous in the honor treatment and lost focus. Since she was in the placebo group, including her in the analyses is actually in favor of our hypothesis.

[15] Note that we have 37 subjects in the Placebo group and 39 subjects in the Testosterone group. The number of subjects required for the one-side Mann-Whitney U test to have a power of 0.8 under the 5% significance level and an expected donation difference of 20% is about 30 for each group.

the hypotheses section, we expect that testosterone has a stronger effect for subjects with high prenatal testosterone exposure (low 2D:4D ratio) compared to subjects with low prenatal testosterone exposure (high 2D:4D ratio). Subfigure (a) shows the result from the honor treatment. Although the treatment effect is higher for the testosterone group for both high and low of the 2D:4D ratio, the difference is slightly larger for the low ratio subjects. This pattern is consistent with our hypothesis. The results are clearer in the shame treatment (subfigure (b)). It illustrates that testosterone only has a positive effect for subjects with lower 2D:4D ratio. We use OLS regression with bootstrap standard errors (999 replications) to test the significance of the interaction effects.[16] The dependent variable is the treatment effect; the independent variables are dummy for testosterone group, dummy for 2D:4D ratio based on median split, and their interaction term. The results are presented in column 1 (honor treatment) and 2 (shame treatment) of table 3.2. Although the signs are consistent with the hypothesis, none of the coefficients reaches significance.[17]

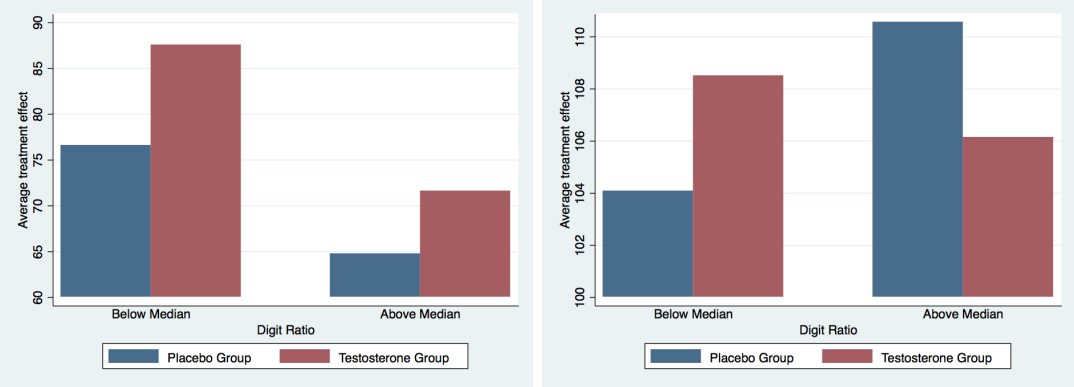
Figure 3.6 presents the results for hypothesis 3. It illustrates the differential effect of testosterone on the treatment effect for subjects with intermediate values of 2D:4D ratio and relatively extreme values of 2D:4D ratio. Brañas-Garza et al. (2013) find that the former group is more altruistic than the latter. Thus, we expect that testosterone has a stronger effect for subjects with relatively extreme values of 2D:4D ratio. The result from the honor treatment is shown in subfigure (a). The difference in treatment effect between the testosterone group and placebo group is indeed higher for subjects with extreme values of 2D:4D. The result from the shame treatment (subfigure (b)) is similar. However, regression results shown in table 3.2 (column 3 is honor treatment and 4 is shame treatment) suggest that the difference is not statistically significant.

The results regarding hypothesis 4 are shown in figure 3.7. The figure demon-

[16] We use bootstrap standard errors due to the small number of observations and we cannot ensure the normality of the data.

[17] Note that it is also possible to separate the first part (baseline condition) and the second part (honor or shame treatment). There are two different methods. The first is using the amount of donation in each condition as the dependent variable instead of using the treatment effect. We did not choose this specification because in this case, we need another dummy variable for the honor and shame condition. This will lead to three way interaction terms, which is hard to interpret. The second method is to run separate regressions for the baseline condition and the honor or shame condition. We performed this exercise, however, most of the results are not significant. The above argument is true for all the regression results presented in table 3.2.

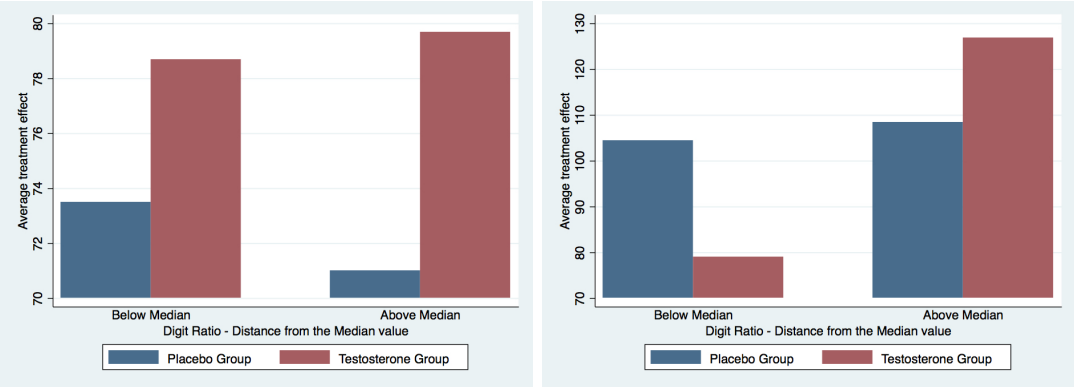
Figure 3.5: 2D:4D Ratio (Digit Ratio)



(a) Honor treatment (b) Shame treatment

Notes: The figure shows the effect of testosterone on the treatment effect, broken down by a median split of the digit ratio (we use the median split to reduce noise of the data).

Figure 3.6: 2D:4D Ratio (Digit Ratio) – Distance from the median



(a) Honor treatment (b) Shame treatment

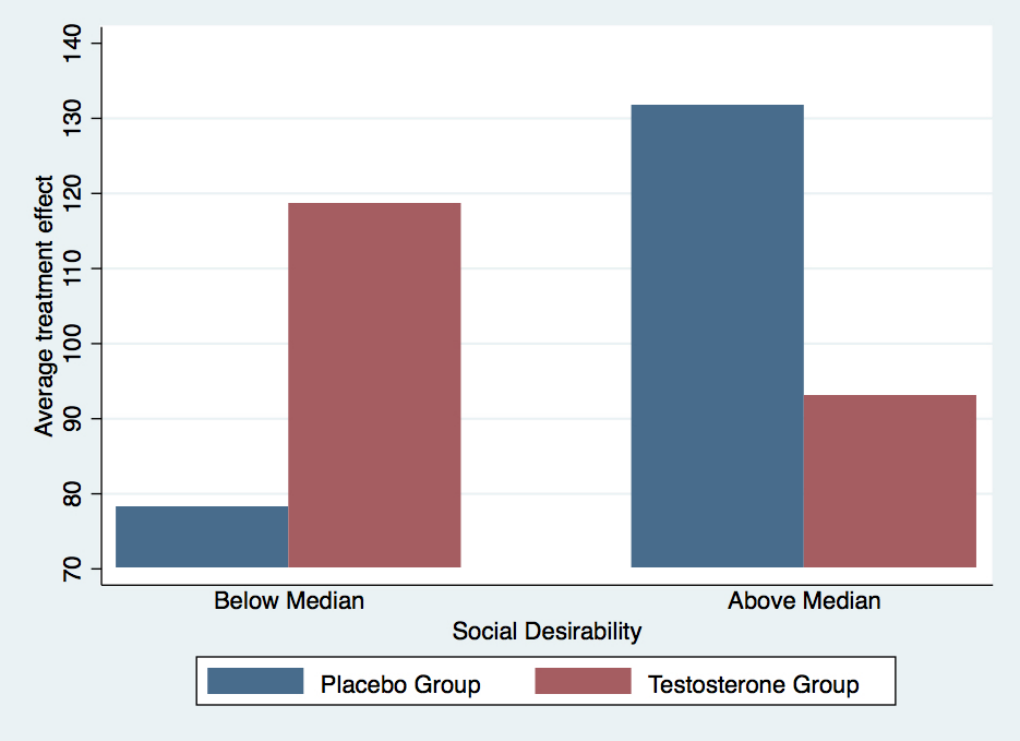
Notes: The figure shows the effect of testosterone on the treatment effect, broken down by a median split of the distance between ones 2D:4D to the median value of the ratio.

strates the effect of testosterone on treatment effect broken down by the social desirability scale.[18] The scale assesses people's tendency to make socially desirable responses to questionnaire questions, which also reflects people's sensitivity to shame. We hypothesize that testosterone has a weaker effect on subjects who have an inherently high level of shame aversion, because these individuals are more motivated by the threat of being publicly 'ashamed' even in the absence of testosterone administration. According to figure 3.7, the effect of testosterone on the treat-

[18]Here, we use the self deception sub-scale of the questionnaire.

ment effect is stronger for subjects who inherently care less about being 'ashamed' and weaker for subjects who are more shame averse. Moreover, according to the regression results in table 3.2 column 5, the testosterone main effect is positively significant and the interaction term is negatively significant, which is congruent with the hypothesis.

Figure 3.7: Social Desirability



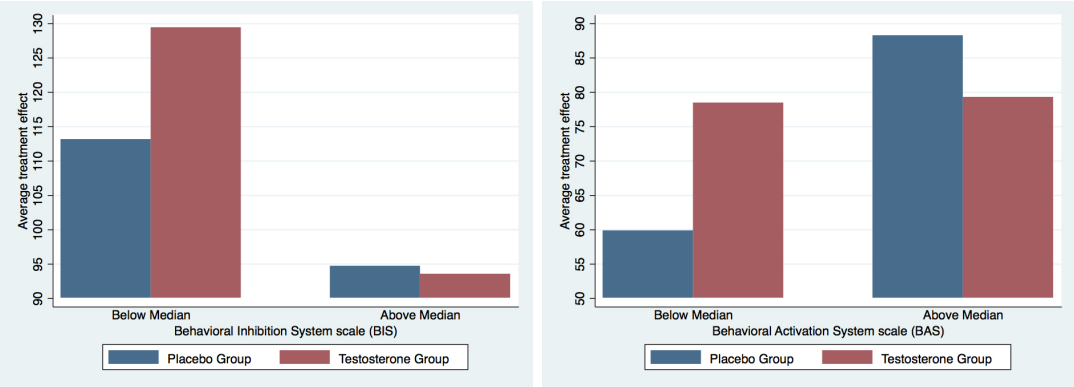
Notes: The figure shows the effect of testosterone on the treatment effect, broken down by a median split of the Social Desirability scale.

Figure 3.8 (a) shows result for hypothesis 5a. Since being publicly identified as the least donator is a thread to one's social image, we expect that subjects who have a more sensitive BIS are less affected by testosterone. This is because high BIS subjects will respond stronger to the potential punishment even without the testosterone administration. In line with the hypothesis, the effect of testosterone on the treatment effect is positive only for low BIS subjects. However, regression results in table 3.2 column 6 suggest that this difference is not significant.

Subjects with a higher BAS score are approach-oriented and respond strongly to potential rewards. Since being appraised as the most generous donator is a reward, the hypothesis is that subjects who have a more sensitive BAS are less

affected by testosterone. The result is illustrated in figure 3.8 (b). According to the figure, the testosterone group's treatment effect is higher than the placebo group only for subjects with lower BAS, which is consistent with the hypothesis. However, regression results in table 3.2 column 7 suggest that this difference is not significant.

Figure 3.8: Behavioral Inhibition System (BIS) and Behavioral Activation System (BAS)



(a) BIS (b) BAS

Notes: The figure shows the effect of testosterone on the treatment effect, broken down by a median split of the BIS/BAS scale.

3.5 Discussion

Our main hypothesis states that testosterone causes individuals to care more about their social image. Therefore, the difference in donation between the honor (shame) and baseline condition should be larger for the testosterone group in comparison to the placebo group. However, our results do not support this hypothesis. There are several factors that might contribute to the null finding, and we discuss each of them in this section.

First of all, the task might be too interesting or enjoyable to perform. Indeed, about 82% subjects stated that they enjoyed a little or enjoyed very much of the real effort task, while only 4 out of 76 subjects stated that they did not enjoy the task. As a consequence, subjects bear less or even no cost in performing the task. Since people have a desire to work on interesting tasks even in the absence of extra incentives (Fehr and Falk, 2002), subjects in our experiment might focus on playing the real effort task instead of the reputation incentives that we intended to

Table 3.2: Regressions for the interaction effects

	Digit Ratio		Digit Ratio from Median		Social Desirability	BIS	BAS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Honor	Shame	Honor	Shame	Shame	Shame	Honor
Testosterone	11.02	4.423	5.167	-25.60	40.52*	16.15	18.64
	(23.23)	(29.54)	(17.85)	(29.09)	(22.98)	(29.80)	(28.02)
Digit Ratio	-11.81	6.469					
	(24.92)	(22.52)					
T x Digit Ratio	-4.141	-8.844					
	(33.04)	(38.32)					
Digit Ratio (From Median)			-2.500	3.812			
			(40.55)	(28.23)			
T x Digit Ratio (From Median)			3.500	44.22			
			(45.73)	(34.53)			
Social Desirability					53.45***		
					(17.36)		
T x Social Desirability					-78.97***		
					(29.20)		
BIS						-18.44	
						(21.79)	
T x BIS						-17.43	
						(36.11)	
BAS							28.31
							(20.14)
T x BAS							-27.54
							(31.31)
Constant	76.56***	104.1***	73.50***	104.5***	78.09***	113.2***	59.86***
	(13.99)	(11.67)	(10.18)	(27.00)	(12.73)	(12.31)	(19.77)
Observations	28	48	28	48	48	48	28

Notes: The dependent variable is the treatment effect. Testosterone is a dummy variable indicating treatment status. T is abbreviation for testosterone and "T x xxx" are the interaction terms between T and other variables. The others variables are dummies constructed by a median split of the relevant scales in order to reduce noise. OLS regressions. Bootstrap standard errors (999 replications) are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

manipulate. It is interesting to investigate whether using more effortful real effort tasks will yield different results.

Secondly, all of our participants were females, because the 4-hour lag for the single sublingual administration of 0.5mg of testosterone to have a neurophysiological effect is only known for women. It might be the case that females are more likely to conform to the majority (Dwenger et al., 2016; Eagly et al., 1981). In our setting, not getting honored and not getting shamed are both the majority outcome, there-

fore, the aversion of being singled out might be too strong so that there is no more room for the effect of testosterone. In a similar manner, Jones and Linardi (2012) find that females suffer disutility from both negative and positive reputation in the context of charity donations. Future work, wielded with the understanding of the time lag for males, might find significant results.

Last, the null result might due to lack of statistical power. We have 29 subjects in the honor treatment and 48 in the shame treatment. Power analyses discussed in the experiment design section suggest that the shame treatment has the power to detect a large effect but is underpowered to detect a small effect, and the honor treatment is underpowered. The results regarding the differential effect of testosterone on different characteristics (hypotheses 2 - 5) provide evidence supporting this claim, as most of the results are in line with the hypotheses but fail to reach significance.

3.6 Conclusion

In this paper, we investigate whether testosterone increases people's reputation concern's in the context of charitable donation. The results suggest that although subjects put more effort in the honor and shame condition relative to the baseline condition, the difference in the increment is not significant between the testosterone and placebo group.

We also investigate the differential effect of testosterone on reputation concerns based on different characteristics such as the sensitivity to testosterone intake, level of altruism, and inherent level of reputation concerns etc. While most of the sign of the results are congruent with our hypotheses, they fail to reach significance at any conventional level.

We believe that whether testosterone leads to reputation based pro-social behavior is an interesting open question. Future work should recruit male subjects, conduct more effortful tasks, and recruit more participants to fill this gap. We also believe that whether the effect of testosterone on reputation concerns differ across gender is an exciting area for future research.

Chapter 4

The Virtue of Honor and the Power of Shame – an Experiment on Charitable Donations

4.1 Introduction

In 2013, there were nearly one million registered charity organizations in the US.[1] They provide aids and services in improving many aspects of our lives and make vital contributions to the well-functioning of our societies. Private donation is the main funding source for charities to keep providing these valuable services. But why do people give? Why do people work hard to earn money but give part of it to unrelated individuals? Andreoni (2006) argues that philanthropy is one of the greatest puzzles in economics.

This paper contributes to the literature that tries to discover the various motivations underlying giving behavior. In particular, we aim to investigate whether it is the pursuit of honor, avoidance of shame, or the combination of both that underly people's charitable behavior. Understanding the effect of the two different motives is important. First, from the donators perspective, if the main reason to donate or donate a higher amount than intended is to avoid the social pressure or the feeling of shame of being the least donator, then the common practice of ranking donators might make people worse off.[2] Second, whether the practice of shaming can

[1] From National Center for Charitable statistics Non-Profit sector in Brief 2015.

[2] DellaVigna et al. (2012) find that there is considerable social pressure cost to turn down a

foster pro-social behavior has important implications for policy makers. Previous studies have discovered that providing monetary incentives to motivate pro-social activities backfires in the case of blood donation (Titmuss, 1970; Mellström and Johannesson, 2008) and working for charity organizations (Gneezy and Rustichini, 2000; Ariely et al., 2009). This is because the provision of extrinsic incentives to pro-social behavior might dilute the intrinsic or reputation value of that behavior (Bénabou and Tirole, 2006). Based on the findings in the psychology literature that not only moral transgressions but also non-moral activities can induce the feeling of shame (Tangney et al., 1996), shaming 'non-compliers' might be an effective way to boost pro-social behavior. Lastly, understand the effect of the two motives is important for charitable organizations to design more efficient fundraising campaigns. If the main motivation to donate is the pursuit of honor, charitable organizations should only public very generous donators. On the other hand, if the main drive is the avoidance of shame, charitable organizations should design indirect and subtle campaigns to utlilize this fact. Nonetheless, this does not imply that charity organizations should 'shame' individuals who donate less as charities face competition in attracting donators in the first place.

In our experimental setting, subjects participated in a real effort task to generate money for charity organizations. The effects of honor and shame were distinguished in three experimental conditions. In the honor condition, the top two donators were identified to other subjects while others' decisions were kept in private. Similarly, in the shame condition, only the least two donators were identified. Subjects' behaviors in these two conditions were compared to the baseline condition in which every subject's donation was private information. There was also a self condition, in which subjects could make money for themselves.[3] We designed a novel ceremony to publicly identify the subjects who donated the most and the least. The ceremony takes advantage of the common procedure of paying experimental subjects. After subjects finished the real effort task as well as a post experimental questionnaire, we asked all of them to come in front of the lab with their signed receipts to wait for their payments. Subjects who were honored or shamed hold yellow receipts and had their payment in yellow envelopes, while other subjects' receipts and envelopes were in white. To make sure everyone notice who was identified, we required subjects to

solicitor.

[3]See the experimental design section for a discussion of why we include the self condition.

remain in the front of the lab until everyone received their payments.

Results from non-parametric tests and regressions indicate that donations made in the honor and shame condition are about 4% and 3% higher, respectively, than the baseline condition. These results imply that the pursuit of honor and avoidance of shame are both important motivations underlying charitable behavior. We also investigate the differential effect of reputation incentives based on subjects' characteristics. The results indicate that subjects who care about their social status, measured by the concern for reputation scale (De Cremer and Tyler, 2005), put more effort in honor and shame conditions relative to the baseline condition. Participants who dislike negative judgments from others, measured by the fear of negative evaluation scale (Leary, 1983), put more effort in the shame condition. Moreover, subjects who were better at the real effort task, measured by their performance in the self condition, put more effort in the honor condition to compete for the 'price'. Last, our main results still hold if we transform the data structure into a between subjects design. This is accomplished by dropping the data from the second condition in each session. Since the baseline and honor (shame) conditions are always the first two conditions and their order were counter balanced, dropping the second condition makes the experimental setting essentially equivalent to a between subjects design.[4] The fact that our results are robust under the between subjects specification suggests that the findings are less likely been driven by the experimenter demanding effect (Zizzo, 2010).

The present paper contributes to the large literature on analyzing the underlying motivations of charitable behavior (see Andreoni (2006) and Vesterlund (2006) for reviews). Researchers have offered several models trying to explain this puzzling behavior. In the early literature, economists assume that charitable giving is motivated by pure altruism (Schwartz, 1970; Young, 1982; Warr, 1982; Roberts, 1984). However, applying this model to analyze charitable behavior yields predictions that are vastly different to what we observe in reality (Andreoni, 1988).[5] In order to better explain the underlying motives of charitable giving, Andreoni (1990) proposed the "warm-glow" model, in which people do not only care for the total provision of the public good but also gain intrinsic value if the donation is made by themselves.

[4]The self condition was always the third and last condition.

[5]The pure altruism model predicts that first, government contribution will crowd out private contributions dollar-for-dollar, only the richest will donate as the economy grows, and the average donation will diminish to zero.

Recent theories, taking into the account the fact that most charity organizations often publicly acknowledge donors, propose that people donate because it grants them prestige or reputation (Harbaugh, 1998). According to the model, prestige might enter people's utility function in two different ways. One is intrinsic, which means that people simply enjoy the feeling of gaining social approval or social status. The other is instrumental: being famous might increases one's income or business opportunities. In a similar model, Glazer and Konrad (1996) proposes that donation is a signaling device to boast ones wealth. In this case, donation is equivalent to conspicuous consumptions in that they demonstrate ones social status (Moav and Neeman, 2010; Charles et al., 2009).

Evidence from numerous lab and field experiments is consistent with the prestige model (see Andreoni and Petrie (2004); Rege and Telle (2004); Van Vugt and Hardy (2010); Ariely et al. (2009); Kataria and Regner (2015); Karlan and McConnell (2014); Reinstein and Riener (2012) for lab experiments and Soetevent (2005); Alpízar et al. (2008); Alpízar and Martinsson (2013); List et al. (2004) for field experiments.) The general finding is that in the 'public' treatment, in which each subject's donation or contribution is made public to other participants, the amount of donation or contribution is significantly higher than the baseline treatment, in which all decisions are kept private.[6] One caveat of this design is that subjects who do not care for gaining social status but dislike social disapproval (being the person who donated the least) also have the motive to increase their donation in the 'public' treatment, since participant's donation or contribution is revealed to others regardless of the amount. Consequently, it is unclear whether it is the pursuit of honor, avoidance of shame, or the combination of both that drive the results.

The rest of the paper is organized as follows. Section 4.2 is the experimental

[6]One notable exception is Dufwenberg and Muren (2006), who find that subjects give less in a dictator game in which their decisions are made publicly in front of other participants. However, the paper differs from the donation literature and our paper in several aspects. One important difference is that the recipients of the dictator game in Dufwenberg and Muren (2006) were the dictator's classmates, whereas the recipients in the donation experiments are usually charitable organizations. It is reasonable to assume that the social ties are stronger among classmates and this might explain why they find different results. Moreover, the dictators were informed about the gender of the recipients, which might also influence their behavior. In the case of charitable donations, on the other hand, gender of the recipients usually does not play an role since the donations are made to charitable organizations. And last, the money allocated to the recipients has direct consequence to dictator's earning in the dictator game, whereas in our experiment, all money are donated to charity.

design. Section 4.3 presents the results and section 4.4 concludes.

4.2 Experimental Design

We conducted 12 sessions in total in March 2015 at the Royal Holloway, University of London Experimental lab (ExpReSS Lab). A total of 133 students were recruited via ORSEE from the subject pool. Subjects came from various disciplines at Royal Holloway. The study was performed in accordance with the Concordat to Support Research Integrity and was approved by Royal Holloway Research Ethnics Committee.

Before each session starts, subjects were informed that they might be publicly identified to other lab participants based on their decisions in the study. They were also informed that their names would be listed outside the experimental lab. We emphasized that if they felt uncomfortable about this procedure, they can receive their show-up fee and leave. All subjects were comfortable to proceed and no one left during the experiment.

The real effort task was programmed in Matlab and the questionnaire was issued using z-Tree (Fischbacher, 2007). All the instructions were written in neutral words to avoid framing affects.[7]

4.2.1 The real effort task

In the experiment, subjects participated in a real effort task to earn money for charity. We opt to use a real effort task instead of a monetary windfall because we believe that the real effort task is more similar to real life situations, in which individuals need to work to earn money and donate part of their earnings to charity. Additionally, studies have documented that subjects behave more selfishly when their endowment was earned via a real effort task compared to windfall money in dictator games (Cherry et al., 2002) and in the context of charitable donations (Carlsson et al., 2013).

The real effort task is based on Kroemer et al. (2014). It consists 12 periods. In each period, subjects are given 30 seconds to press the 'LEFT' key of the computer

[7]See Appendix A.2 for the experimental instructions for the honor treatment. The instructions for the shame treatment is very similar and is available upon request.

mouse as fast as possible. After each period, subjects need to hit the 'SPACE' bar to continue to the next. We choose this task for several reasons. First, it does not require any prior knowledge. Second, it is very simple and hence there is little learning possibility. Third, the task is boring and effortful, therefore costly to perform. And last, the output of the task has no value to the experimenter. This rules out the possibility that subjects perform the task in order to reciprocate the payments made by the experimenter.

We employ four different payoff levels: each click may worth 1, 3, 6, or 10 points.[8] There are three rounds for each payoff level and the order is randomly decided for each subject. Before each period starts, information about the payoff level of the current period will appear on subjects' computer screen (see figure 4.1). Additionally, to prevent subjects from loosing time due to distracted attention, a count down from three to one is launched before each period starts. Since clicking generates a lot of noise, subject's effort might be affected by others due to peer effects or conformity. To control for this, we ask subjects to put on a headset and listen to soft music. This design ensures that subjects can progress at their own pace without receiving disturbance from other participants. All the points subjects earned in the real effort task are exchanged to British Pound Sterling at rate: 100 points = 2 Cents.

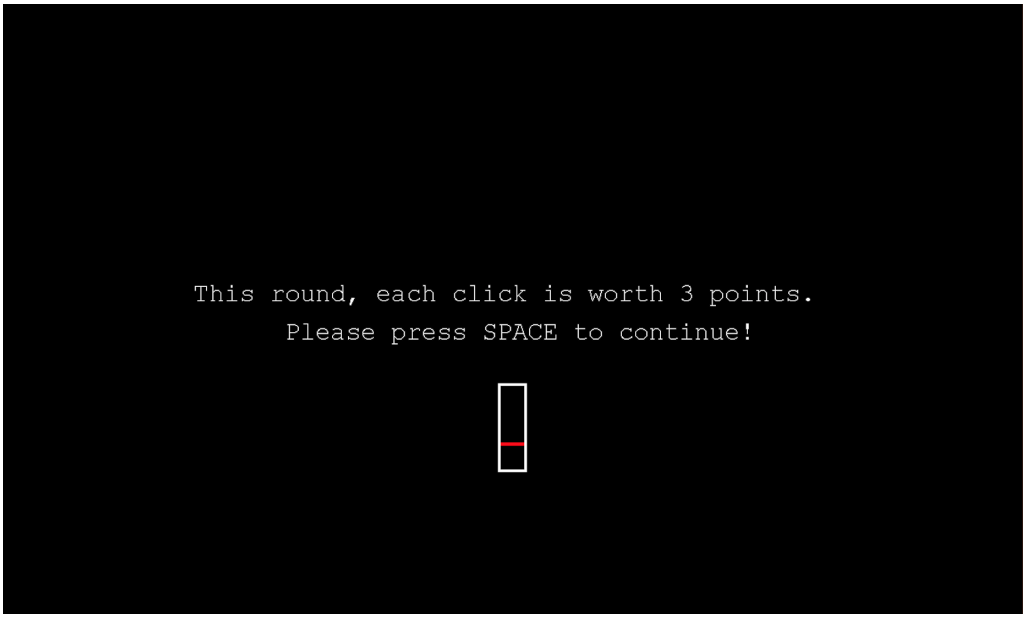
4.2.2 Experimental Conditions and Treatments

There are a total of four experimental conditions in our experimental setting, namely, baseline condition, shame condition, honor condition, and self condition. Subjects cannot earn money for themselves in baseline, honor, and shame conditions. All money are donated to charity. This design ensures that subject's donation is driven by altruistic or reputation concerns, not monetary incentives. Additionally, the presence of monetary incentives might crowd out donation because observers are not sure whether the donation is for good deeds or for money (Bénabou and Tirole, 2006; Ariely et al., 2009). Since all money are donated to charity in these conditions, we refer to these conditions as donation conditions.

In the baseline condition, the amount of donation is kept in private. In this case,

[8]This design allows us to investigate whether subject's behavior in the real effort task is rational or not. Since effort is costly, a efficient strategy is to put less effort in lower payoff levels. This is indeed what we find (see figure A.1 in the Appendix A.1)

Figure 4.1: Information screen before each period starts



reputation concerns do not have any effect on the amount of donation. Warm-glow (Andreoni, 1990) and sympathy towards the recipients are the two main reasons to donate in the baseline condition. In the honor condition, only the top two donators are publicly identified, while other participants remain anonymous. Please note that we only reveal the identity of the two top donators.[9] Neither the amount of donation nor their chosen charity is revealed. Individuals who care about social status or prestige have extra incentives to make higher donations in order to be 'honored'. On the other hand, subjects who are not motivated by gaining social status do not need to increase their donations to avoid social disapproval because their donations are not publicized. The shame condition is the same as the honor condition, with the exception that the least two donators are publicly identified. In the shame condition, the aversion of shame or social pressure is the extra motivation for people to donate more than the baseline condition. These three conditions allow us to isolate the effect of honor and shame on charitable behavior.

Subjects can earn money for themselves in the self condition and the amount they earn is private information. We implement the self condition for two reasons. First, the money they earn in the self condition can serve as an indicator for their ability in performing the task under the rather general assumption that subjects will

[9]There were about 11 subjects in each session, for both the honor and shame treatments.

try to earn as much as they can.[10] Second, when subject's donation in the honor condition is similar to the shame condition, their performance in the self condition is important to distinguish whether they reached a physical limit or the effect of shame and honor is indeed similar.

We implement a within–between subjects 2x2 design. The honor treatments consist baseline, honor, and self conditions (note that each condition consists of 12 periods of clicking.) The shame treatments consist baseline, shame, and self condition. The within element of our design is that subjects participate in both baseline and honor (or shame) conditions in each session. To control for order effects, the sequence of the baseline and honor (shame) conditions are counter balanced. Note that the self condition is always the last condition. The between element is that subjects only perform one of the honor and shame treatments. The four experimental treatments are summarized in table 4.1.

Table 4.1: Summary of Experimental Treatments

	Normal Order	Reverse Order
Honor Treatment	Baseline – Honor – Self (34)	Honor – Baseline – Self (32)
Shame Treatment	Baseline – Shame – Self (36)	Shame – Baseline – Self (31)

Notes: The numbers in the parentheses are number of subjects in each treatment.

4.2.3 Honoring and Shaming ceremony

We designed a novel honoring and shaming ceremony, which we believe is less artificial than methods used in other papers, such as showing participants' contributions and their names (and photos) on computer screens (Andreoni and Petrie, 2004; Samek and Sheremeta, 2014), discussing their donations in front of other participants (Ariely et al., 2009), and asking participants to write their name on the blackboard under the word "I donated the most (least)" (Jacquet et al., 2011).

The ceremony takes advantage of the common procedure of paying experimental subjects. After subjects finish all the conditions as well as a post-experimental questionnaire, we ask all of them to come to the front of the lab with their signed receipt and await their payments. Subjects who are honored or shamed hold yellow

[10]This is very likely the case in our setting because subjects can only make money for themselves. More importantly, subject's decision does not affect other subjects, therefore, social preferences do not play a role.

receipts and have their payment in yellow envelopes, while others' receipts and envelopes are in white.

To make sure everyone notices which subjects are identified, we ask subjects to remain in the front of the lab until everyone receives their payment. Because all the subjects are required to come to the front of the lab and stand in front of others, this design can control for the 'exposure aversion' effect, in which some subjects dislike being exposed in front of others. Jones and Linardi (2014) find that there is a type of individuals that possess 'Wallflower' preferences. Subjects who have this preference tend to avoid being honored as well as being shamed, in other words, they dislike any kind of public exposure. The presence of this type of preference will underestimate the effect of honor and overestimate the effect of shame. Subjects are free to leave after all of them get their payments.

In order to make the honoring and shaming more formal and credible, we also inform subjects that we will make a public list of participants and post the list outside the Royal Holloway experimental Lab. The name of those who get honored or shamed in each session will be highlighted in yellow.

4.2.4 Experimental Procedure

After subjects arrived at the lab and agreed to proceed, we conducted a 'meet and greet' stage in order to create some social connections amongst subjects. Subjects who sat in the same row were asked to chat with each other. To help them 'break the ice', the experimenter announced in public that they could exchange their names, field of study, where they came from, and their hobbies. The 'meet and greet' stage lasted about five minutes. Any communication between subjects was prohibited after this stage. Afterwards, the instructions for the first condition were handed out to the subjects. They needed to answer the control questions correctly in order to proceed. Moreover, subjects received a list of ten charitable organizations with descriptions. This list incorporated popular charities and not-for-profit organizations in the UK. Subjects were asked to choose one preferable organization. Before the actual task began, subjects had four practice rounds, one for each pay level, to familiarize themselves with the real effort task. Participants knew that the session consists of several parts, but they did not know the content of the future parts before the corresponding instructions were provided.

The honoring and shaming stage was the very last part of the experiment, after they finished all three conditions and a questionnaire. No feedback about their performance was given until they completed all the conditions and the post experimental questionnaire. This design ensures that subject's behavior in the second condition is not affected by the outcome of the first condition.

4.3 Results

4.3.1 Identification

Since subjects participated in both the baseline and honor (or shame) condition in each session, the dependent variable of interest is the treatment effect, defined as the amount of donation made in the honor or shame condition minus the amount of donation made in the baseline condition. The treatment effect is able to identify subjects' reputation concerns, while at the same time control for a number of individual specific unobservables.

Suppose that subject's donation in the baseline condition is determined by the following equation:

$$\text{Donation}_i^{Baseline} = A_i + W_i + u_i \tag{4.1}$$

where A_i denotes subject i's ability in performing the task, W_i represents i's intrinsic value of donation (warm glow), and u_i is other unobservables that might affect i's donation. In the honor or shame condition, everything remains the same except that reputation concerns might also affect their behavior. This can be modeled as:

$$\text{Donation}_i^{Honor\ or\ Shame} = A_i + W_i + R_i + u_i \tag{4.2}$$

where R_i denotes i's reputation concern. The treatment effect is calculated from subbstracting equation 4.1 from equation 4.2. The unobservables such as ability and intrinsic motivation are canceled out, and the remaining term, R_i, represents subject i's reputation concern.[11]

[11]The order of the conditions might affect our results. In particular, subject's behavior in the baseline condition when the order is baseline – honor (or shame) might be different from their behavior when the order is reversed. We discuss this in detail in the robustness check subsection and show that order has minimal affect on subject's behavior .

4.3.2 The Effect of Honor and Shame

Table 4.2 presents the average amount of money subjects made in each condition. The unit is in British Pence. Column (1) and (2) shows the result for the honor treatments and shame treatments respectively.[12] The treatment effect is significantly higher than zero in both the honor and shame treatments (Wilcoxon signed-rank test, z = 3.312, p = 0.0009, N = 66 for honor treatments; Wilcoxon signed-rank test, z = 2.511, p = 0.0120, N = 67 for the shame treatments). Specifically, subject's donation in the honor condition is about 4.1% higher than the baseline condition. The difference is 3.5% in the shame treatments.

Moreover, donation in the honor and shame conditions is not significantly different from each other (Mann-Whitney U test, z = 0.506, p = 0.6127, N = 133). The similar amount of donation in the honor and shame conditions is not due to the fact that subjects reach their physical limit in either conditions, as demonstrated by the fact that subjects perform even better in the self condition. Their performance in self condition is 3.9% higher than the honor condition (Wilcoxon signed-rank test, z = 4.162, p = 0.0000, N = 66) and is 6.1% higher than the shame condition (Wilcoxon signed-rank test, z = 4.348, p = 0.0000, N = 67).

Table 4.2: Average amount of money made in each condition

	(1) Honor treatments		(2) Shame treatments	
Baseline condition	229.3	(45.7)	229.5	(47.5)
Honor condition	238.6	(48.9)	-	-
Shame condition	-	-	237.5	(44.9)
Self condition	247.8	(51.0)	251.9	(43.8)
N	66		67	

Notes: The variable is the mount of money subjects generated in Baseline, Honor, Shame, and Self conditions. The unit is British Pence. Standard errors are in the parentheses.

We also investigate the proportion of subjects who achieve a positive treatment effect. If subjects in our experiment are not motivated by reputation concerns but click randomly instead, the proportion of subjects who make a positive treatment effect should not be significantly different from 50%. This is not the case. We find

[12] We separately report the baseline donation in the honor and shame treatments as subjects participated in both baseline and honor (shame) conditions and the statistical tests are based on paired tests.

that more subjects exert a positive treatment effect in both the honor and shame treatments. In the honor treatments, 68.18% of the subjects achieve a positive treatment effect, which is significantly higher than 50% (Binomial test, p = 0.0043, N = 66). The proportion in the shame treatments is 70.1% and is also higher than 50% (Binomial test, p = 0.0007, N = 67). Additionally, we find that the proportion is similar in honor and shame treatments (Two-sample test of proportions, z = 0.2457, p = 0.8059, N = 133.)

4.3.3 The Effect of Reputation Concerns on Different Characteristics.

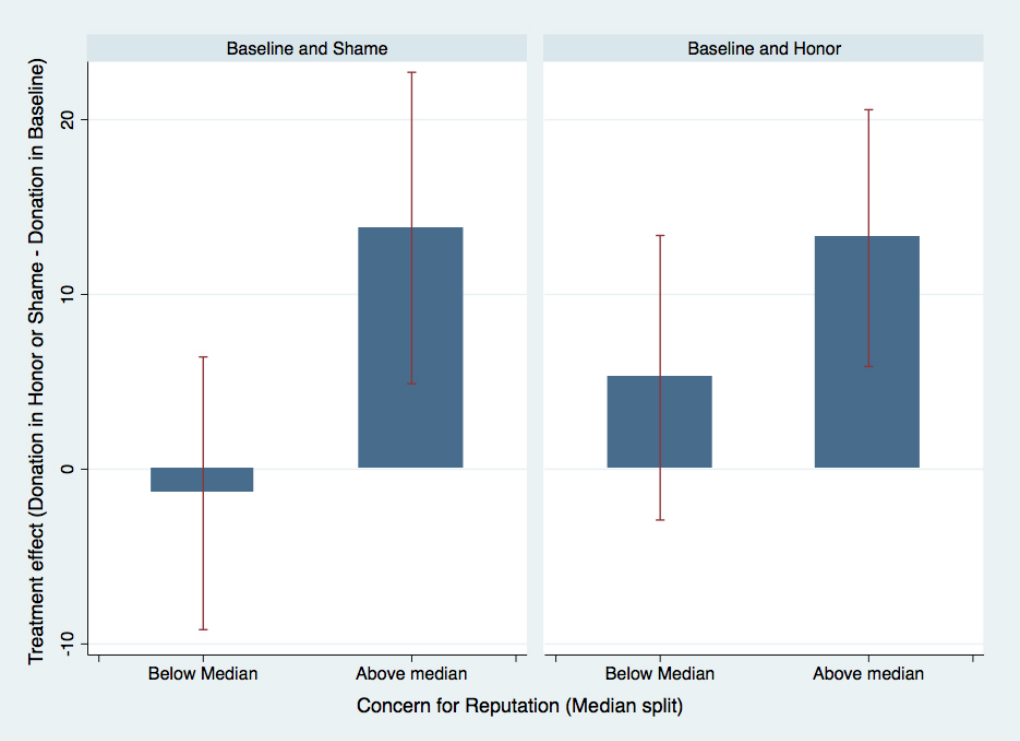
We have shown that subjects respond to image incentives by increasing their effort in the honor and shame conditions relative to the baseline condition. But who are more affected by the reputation incentives? In this subsection, we show that characteristics such as concern for reputation, fear of negative evaluation, and ability in performing the task have profound influences on subject's behavior.

Subjects who value their social status relatively more than others should try harder in the honor and shame conditions to gain status and to avoid the lost of reputation, respectively. We test this hypothesis using the 7-item Concern for Reputation Scale (De Cremer and Tyler, 2005), which was administrated after the real effort tasks and before the honoring and shaming ceremony. We find that treatment effect and the Concern for Reputation Scale is positively correlated in both the honor and shame treatments (Spearman Correlation, Rho = 0.2818, p = 0.0219 for honor treatments and Rho = 0.2393, p = 0.0512 for shame treatments.) This means subjects who care more about social status increase their donation more from the baseline condition to the honor or shame condition compared to subjects who care relatively less about their social image. We further break down the sample by a median split of the scale and investigate the treatment effect for each sub-group. The results are shown in figure 4.2. The left panel is the data from the shame treatment.[13] For subjects who relatively care less about reputation

[13]For illustration purpose, we exclude two outliers in this bar graph. These two outliers were both in the shame treatment. One of them had a very high treatment effect while the other had a very low treatment effect. Including these two subjects will make the 95% interval very wide but does *not* affect the mean. Please note that the two outliers only affects the appearance of the graph, we therefore include them in the following statistical tests.

(reputation scale below median), the treatment effect is not significantly different from zero (two sided t-test, t = 0.0417, p = 0.9670, N = 27).[14] On the other hand, for subjects who care more about reputation (scale is higher than median), the treatment effect is significantly higher than zero (two sided t-test, t = 3.1313, p = 0.0033, N = 40). Results from the honor treatments (right panel) are similar. Treatment effect is not significantly different from zero for subjects whose Concern for Reputation Scale is below median (two sided t-test, t = 1.3110, p = 0.1995, N = 32). For subjects whose Concern for Reputation Scale is above median, the treatment effect is significantly larger than zero (two sided t-test, t = 3.6613, p = 0.0009, N = 34).

Figure 4.2: Treatment effect by median split of reputation concern scale



Notes: Treatment effect is defined as the amount of donation made in honor or shame condition minus the amount of donation made in baseline condition. Below and above median implies whether subject's score from the concern for reputation scale is below or above median. For illustration purpose, we exclude two outliers in this bar graph. These two outliers are both in the shame treatment. One of them had a very high treatment effect while the other had a very low treatment effect. Including these two subjects will make the 95% interval very wide but does *not* affect the average. Please note that the two outliers only affects the appearance of the graph, we therefore include them in the statistical tests.

[14] We use t-test here because 95% confidence intervals are used in the bar graph. Please note that the results are the same if we use the Wilcoxon signed-rank test.

Another natural hypothesis is that subjects who dislike receiving negative evaluation should try harder in the shame condition to avoid being the least donator. Using the Fear of Negative Evaluation scale (Leary, 1983), we show that the results are consistent with the hypothesis. Subjects' scores on the scale are positively correlated with the treatment effect for the shame treatment (Spearman Correlation, Rho = 0.2214, p = 0.0718).[15] The results are the same if we divide the subjects by a median split of the scale (figure 4.3).[16] According to the bar graph, subjects who dislike negative evaluation have a treatment effect significantly larger than zero (two sided t-test, t = 2.5205, p = 0.0157, N = 41). The treatment effect is not statistically different from zero for subjects whose scale is below median (two sided t-test, t = 0.4067, p = 6878, N = 25).[17]

Subject's ability in performing the task also affects their responsiveness to the image incentive. If subjects form correct beliefs about their own ability comparative to other participants, then, in the honor treatments, only subjects who are good at this task have the incentive to compete for the 'price'. Similarly, in the shame treatments, only subjects who are bad at this task should try harder to avoid the 'stigma'. We use subject's performance in the self condition as their ability measure.[18] We find that the treatment effect and ability is positively correlated in the honor treatments (Spearman correlation, Rho=0.2945, p=0.0164, N = 66) but not in the shame treatments (Spearman correlation, Rho = -0.0702, p = 0.5722, N = 67). In the following, we break down the sample by a median split of ability measure and report the average treatment effect for each sub-group. Under this specification, the result from the shame treatment is congruent with the hypothesis. The results are presented in figure 4.4. The left panel is the data from the shame treatment.[19] For subjects who are relatively bad at the task (performance in the self condition below median), the treatment effect is significantly larger than zero (two sided t-test, t = 3.0753, p = 0.0046, N = 30).[20] On the other hand, for subjects who

[15]The correlation is not significant in the honor treatment (Spearman Correlation, Rho = 0.1717, p = 0.1680), which is expected because being honored is not a punishment.

[16]As explained in figure 4.2, we exclude two outliers for illustration purpose but include them in the statistical tests.

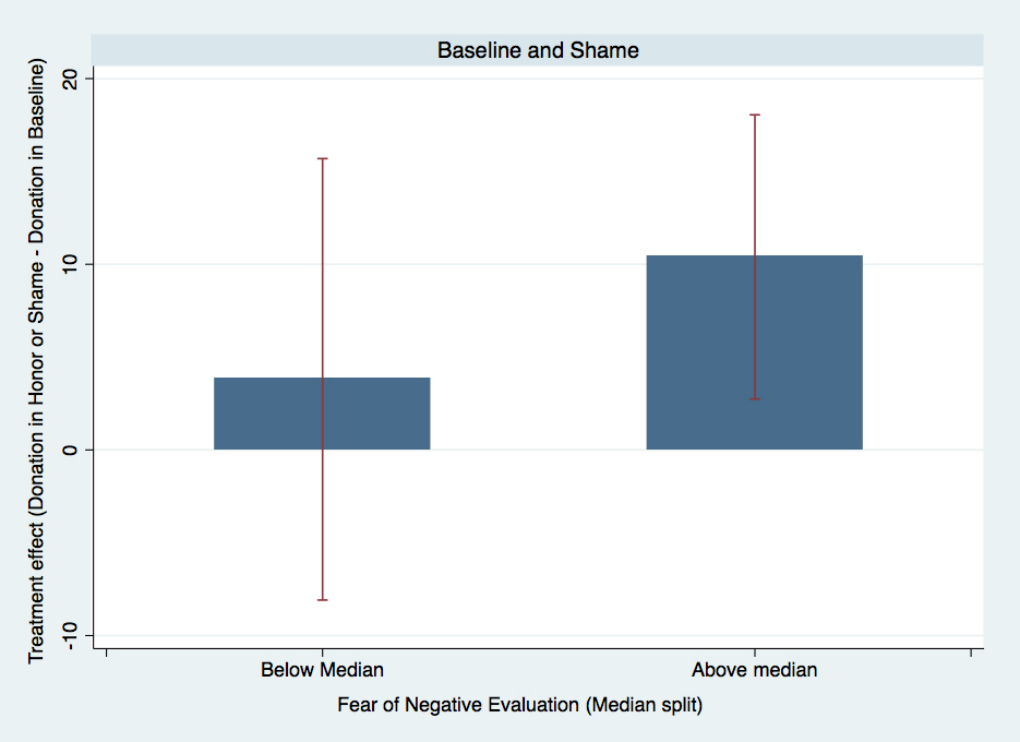
[17]The results are the same if we use the Wilcoxon signed-rank test.

[18]The results are similar if we use the maximum clicks they achieved in the experiment as ability measure.

[19]As in figure 4.2 we exclude two outliers in the graph, but include them for statistical tests.

[20]The results are the same if we use Wilcoxon signed-rank tests.

Figure 4.3: Treatment effect by median split of Fear of negative evaluation scale

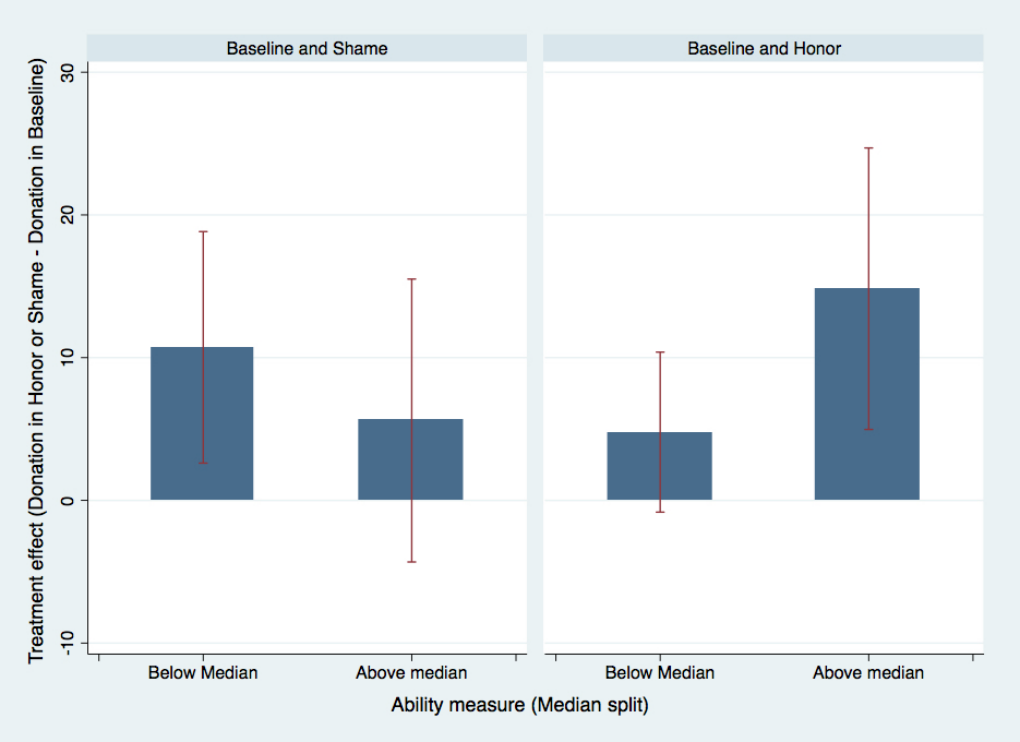


Notes: Treatment effect is defined as the amount of donation made in honor or shame condition minus the amount of donation made in baseline condition. Below and above median implies whether subject's score from the fear of negative evaluation scale is below or above median. As explained in figure 4.2, we exclude two outliers for illustration purpose but include them in the statistical tests.

are more competent (ability measure higher than median), the treatment effect is not significantly different from zero (two sided t-test, t = 0.6575, p = 0.5151, N = 37). Results from the honor treatments (right panel) are the opposite. For subjects whose ability is below median, the treatment effect is weakly higher than zero (two sided t-test, t = 1.7314, p = 0.0922, N = 36). On the other hand, for subjects who are good at this task, the treatment effect is significantly larger than zero (two sided t-test, t = 3.0753, p = 0.0046, N = 30).

The findings regarding the differential effects of reputation concerns on subjects' characteristics are replicated using random effects panel regressions. The results are presented in table 4.5. Column 1 corresponds to the results presented in figure 4.2. Concern for Reputation (Dummy) is a dummy variable based on the equal split of subjects' scores on the Concern for Reputation Scale. The Honor and Shame coefficients are not significant, suggesting that subjects who care relatively less about reputation do not put more effort in the honor and shame conditions in comparison

Figure 4.4: Treatment effect by median split of Ability measure



Notes: Treatment effect is defined as the amount of donation made in honor or shame condition minus the amount of donation made in baseline condition. The ability measure is based on the amount of money they made for themselves in the self condition. As explained in figure 4.2, we exclude two outliers for illustration purpose but include them in the statistical tests.

to the baseline condition. The interaction terms between the scale and the conditional dummies (Honor or Shame X Reputation) are positive but not significant. However, post regression tests suggest that subjects whose score on the concern reputation scale is higher than median contribute significantly more in the honor and shame condition compared to the baseline condition (Chi Square test, $p = 0.013$ for the Honor; $p < 0.01$ for Shame).

Column 2 of table 4.5 replicates the result illustrated in figure 4.3. The Shame is not significant, while the Shame X Fear of Negative Evaluation Dummy is significantly positive. This suggests that subjects who dislike others to hold negative perceptions of themselves tried harder in the shame condition in order to avoid the stigma.

Column 3 mirrors the result presented in figure 4.4. The Honor coefficient is not significant while the Honor X Ability is. This indicates that only subjects who are better at this task actually tried to compete for the "price". For the

shame treatment, the situation is the opposite. The Shame coefficient is significantly positive while its interaction term with Ability is negative (but not significant). This is congruent with the hypothesis that only subjects who are really bad at this task need to try harder in the shame condition to avoid being the least donator.

Figure 4.5: Regressions about the differential effects of reputation concerns on subjects' characteristics

	(1) Reputation	(2) Fear	(3) Ability
Honor Condition	0.435		0.235
	(0.551)		(0.298)
Shame Condition	-0.0379	-0.319	1.090***
	(0.882)	(0.775)	(0.292)
Concern for Reputation (Dummy)	-1.353**		
	(0.684)		
Honor X Reputation	0.567		
	(0.685)		
Shame X Reputation	1.272		
	(0.951)		
Fear of Neg. Evaluation (Dummy)		-1.876*	
		(0.998)	
Honor X Fear			
Shame X Fear		1.579*	
		(0.921)	
Ability (Dummy)			4.559***
			(0.540)
Honor X Ability			1.671***
			(0.637)
Shame X Ability			-1.143
			(0.752)
Constant	19.87***	20.30***	16.82***
	(0.559)	(0.817)	(0.271)
Observations	3192	1608	3192

Notes: Concern for Reputation (Dummy) is a dummy variable which equals one if the subject scored equal to or above the median on the concern for reputation scale and equals zero otherwise. Fear of Negative Evaluation (Dummy) is the median split based on their score on the Fear of Negative Evaluation scale. And the Ability (Dummy) is based on the median split of the money they made in the self condition. Cluster Standard errors (individual level) are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

4.3.4 Robustness Checks

In this section, we first show that our main results are robust to the inclusion of a set of control variables. Second, we present evidence that the order of the conditions has minimal affect on subject's behavior. And last, we replicate the main result in a between subjects design data structure.

The regression results are presented in table 4.3. The dependent variable is the amount of donation made in each round. In column (1), we only include the honor and shame condition dummy as independent variables. The honor coefficient implies that in each round, subjects on average donate 4% more in the honor condition in comparison to the baseline condition, and the difference is significant at 5%. The difference is 3.5% for the shame condition, but it is marginally insignificant ($p = 0.104$). This is due to the two outliers mentioned before in figure 4.2. These two observations have little effect on average donation but inflate the standard errors dramatically. After dropping the two outliers in column (2), the shame coefficient becomes highly significant. We add a set of control variables in column (3). The controls include pay level effects, round effects, gender, age, the number of participants they know, enjoyment of the task, concern for reputation scale, and fear of negative evaluation scale. Additionally, in order to control for order effects, we add a dummy variable, Reverse Order, which equals 1 if the order of the treatment is honor (or shame) – baseline and equals zero otherwise. Both the honor and shame coefficients remain significant.

We also investigate whether subject's behavior in the baseline, honor, and shame condition is affected by the order of the conditions. For example, we compare subject's donation in baseline condition when the order is baseline – honor (shame) to their donation when the order is honor (shame) – baseline. The results indicate that the order has a minimal effect on subject's behavior. For baseline and shame conditions, the difference is not significant at any conventional levels (Mann-Whitney U test, $p > 0.29$ for the baseline conditions in the honor treatment and in the shame treatment; $p > 0.72$ for the shame conditions.) We find that subjects contribute more in the honor condition when it is the second condition and the difference is weakly significant (Mann-Whitney U test, $p = 0.07$). Further analyses suggest that this is driven by session four, which was a honor – baseline treatment. Subjects from session four had lower ability than others, as they performed significantly worse in

all conditions, including the self condition. Consequently, they dragged down the average contribution in the honor condition when the order is honor – baseline. When we exclude session four from the analyses, the difference vanishes (Mann-Whitney U test, $p > 0.72$).

To provide additional evidence that our results are less likely driven by order effects, we show that our main result still holds in a between subjects structure. This is accomplished by discarding the data from the second condition in each session. In this case, subject's behavior in the baseline, honor, and shame conditions is not affected by previous experience. The results are shown in table 4.4. The dependent variable is the amount of clicks in each round divided by the maximum clicks they achieved throughout the experiment;[21] we use this relative measure to control for subjects' ability differences.[22] The results indicate that the amount of clicks in the honor condition is 3.08% closer to their best performance compared to the baseline condition. The result is similar for the shame condition.[23]

4.4 Conclusion

Previous studies have established the importance of reputation or prestige motivations on people's charitable behavior. However, these studies remain silent on which aspect of reputation concerns, namely, pursuit of honor, avoidance of shame, or the combination of both, is the primary underlying motivation. In this study, by conducting a real effort experiment and a novel honoring and shaming ceremony, we aim to disentangle these two effects.

We find that both honor and shame are both important motivations. Donation made in the honor condition is about 4% higher than the baseline condition and the donation made in the shame condition is about 3% higher than the baseline condition. Moreover, donation made in honor and shame conditions are not significantly different from each other, which suggests that two motives are equally

[21] We use clicks instead of donation here because the amount of donation in each round is also affected by the pay level in that round.

[22] We also run a fixed effect regression including the self condition, which is another way the take into account ability heterogeneity. The results are the same.

[23] The Male coefficient is negative and highly significant. This result suggests that males put more effort in the self condition related to the donation conditions compared to females. We also run the regression with honor and shame dummies interact with Gender. The interaction terms are negative but not significant.

Table 4.3: Random effects panel regression

	(1)	(2)	(3)
Honor Condition	0.772**	0.676**	0.777***
	(0.340)	(0.334)	(0.209)
Shame Condition	0.678	0.769**	0.625**
	(0.417)	(0.307)	(0.253)
Reverse Order			-0.591
			(0.587)
Male			2.237***
			(0.635)
Concern for Reputation			-0.817*
			(0.484)
Fear			0.523
			(0.389)
Enjoy Task			0.868***
			(0.199)
Pay Level			4.055***
			(0.0680)
No. of Subjects they Know			-0.288
			(0.244)
Age			-1.504**
			(0.628)
Age Squared			0.0307***
			(0.0117)
Round			0.114*
			(0.0590)
Round Squared			-0.00251
			(0.00176)
Constant	19.11***	19.23***	13.73
	(0.336)	(0.322)	(8.356)
Observations	3192	3144	3144

Notes: The dependent variable is the amount of donation made in each period. Honor condition and Shame condition are dummy variables. Reverse Order is a dummy variable which equals 1 if the Order of the treatment is Honor (Shame) – Baseline. Fear refers to the fear of negative evaluation scale. In column 2 we dropped the two outliers mentioned before (figure 4.2. Cluster Standard errors (individual level) are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

effective in attracting donations. We also find that subject's concern for reputation, fear of negative evaluation, and ability in performing the real effort task predict

Table 4.4: Random effects panel regression - Between subjects data structure

	(1)	
Honor Condition	0.0307*	(0.0185)
Shame Condition	0.0303*	(0.0170)
Male	-0.0783***	(0.0179)
Concern for Reputation	-0.00118	(0.0120)
Enjoy Task	0.00327	(0.00548)
Fear of Neg. Evaluation	-0.0189*	(0.0112)
Pay Level	0.00695***	(0.000783)
No. of Subjects they Know	-0.000709	(0.00511)
Age	-0.0104	(0.0198)
Age Squared	0.000156	(0.000371)
Round	-0.0321***	(0.00350)
Round Squared	0.00141***	(0.000163)
Constant	1.151***	(0.262)
Observations	1596	

Notes: The dependent variable is the amount of clicks in each round divided by the maximum clicks they achieved in one round throughout the experiment. Honor condition and Shame condition are dummy variables. Cluster Standard errors (individual level) are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

their responsiveness to the reputation incentives.

There are several open questions for future research. First is to investigate the effect of shame on long term relationships between charity organizations and their 'customers'. In particular, does the practice of shaming individuals who donate less than others make organizations worse off in the long run? Second, there are a number of important motivations underlying charitable behavior, namely, sympathy of the recipients, warm-glow, pursuit of honor, and aversion of shame. It would be interesting to investigate the relative importance of these motivations. Last, it is interesting to test whether shaming can encourage pro-social behavior in other contexts, such as blood donation, purchase of low emission vehicles, and volunteering.

Appendix A

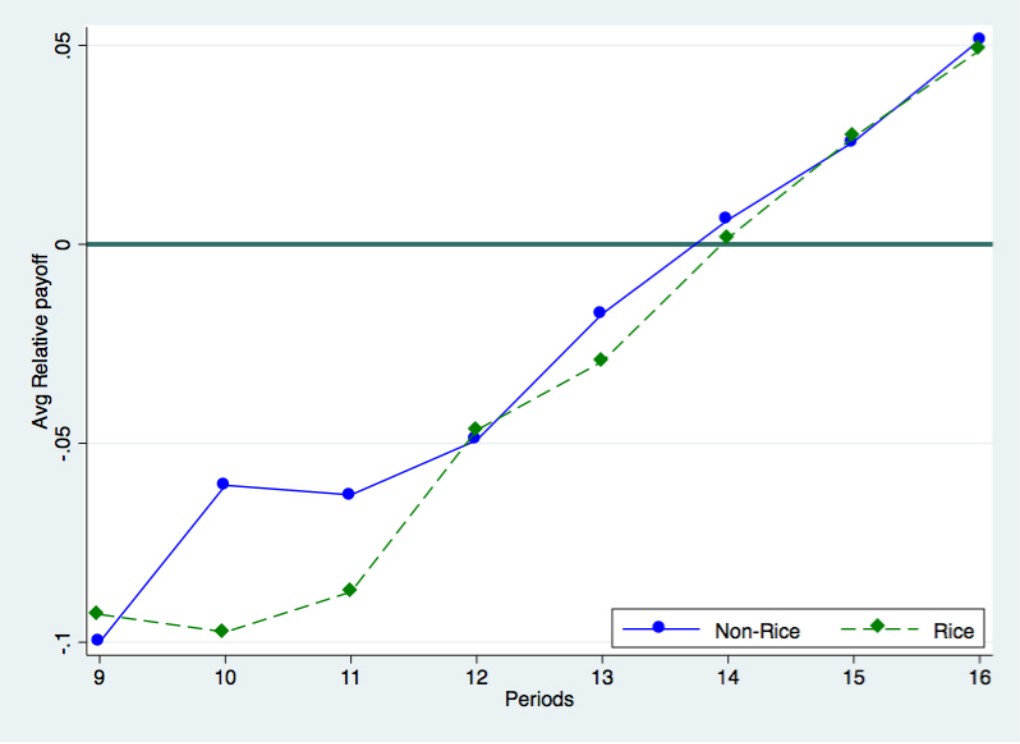
Appendices

A Appendix to Chapter 2

A.1 Relative Welfare

The relative welfare in the PGG with punishment is presented in figure A.1. The relative welfare is constructed as the following. We first compute the average payoff in each period of the no-punishment condition. This is done separately for rice and non-rice subjects. We then subtract this average from individual's payoff in the punishment condition. Finally, the difference is normalized by the average payoff in the no-punishment condition. This yields the evolution of payoff in the punishment condition relative to the no-punishment condition.[1]

Figure A.1: Evolution of relative welfare



Notes: This figure presents the relative welfare in the punishment condition with respect to the no punishment condition. This is constructed by first calculating the welfare difference between the two conditions and normalize the difference by the welfare in the no punishment condition.

In the first period of the punishment condition, rice and non-rice subjects earn

[1]Fehr and Gächter (2000) uses the group level average payoff to calculate the relative payoff. We cannot compare group level difference between the two conditions because subjects randomly regroup after the no-punishment condition.

about 10% less compared to the first period of the no punishment condition. The difference becomes smaller over time and subjects in both regions start to make a relative payoff gain from period fourteen onwards. Taking all periods together, the average relative payoff is -2.6% in non-rice and -3.5% in rice. The difference is not significant at any conventional level (Mann-Whitney U test based on group level averages: $p > 0.72$). Moreover, in periods 10 and 11, in which the difference between rice and non-rice is the largest, the differences are also not significant (Mann-Whitney U test based on group level averages: $p > 0.25$ for both periods).

A.2 Additional Regression Analyses

Table A.1: Tobit Regressions Regarding the Effect of the Percentage of Rice Paddy fields on Punishment Behavior among Rice Farming Regions.

	Rice Regions		Non Rice Regions	
	(1)	(2)	(3)	(4)
Perc. Paddy Field (Municipality)	0.0691**	0.0625**	-0.0182	-0.0173
	(0.0311)	(0.0315)	(0.0265)	(0.0227)
Anti-Social Punishment	-0.591	-0.197	-1.381***	-0.941***
	(3.854)	(3.219)	(0.232)	(0.258)
Perc. Rice Paddy X Anti-Social Punishment	-0.0167	-0.0122	0.0170	0.00216
	(0.0468)	(0.0410)	(0.0347)	(0.0319)
Punisher Contribution		0.0517		0.0619
		(0.0648)		(0.0495)
Punished Contribution		-0.0659**		-0.0299
		(0.0308)		(0.0250)
Other Two Member Avg Contribution		-0.119**		-0.0778**
		(0.0601)		(0.0355)
GDP per cap. (Municipality)		0.00657		-0.0784
		(0.0684)		(0.0578)
Holistic Thinking		-1.247		-1.417**
		(0.841)		(0.615)
From Rural (Dummy)		0.167		0.151
		(0.387)		(0.353)
Relative Income		0.375		0.0240
		(0.279)		(0.292)
Risk Attitude		-0.0464		0.213*
		(0.144)		(0.121)
Male		1.023*		-0.212
		(0.536)		(0.279)
Natural Science (Dummy)		-0.0197		-0.174
		(0.330)		(0.374)
Priming (Dummy)		0.314		0.878**
		(0.573)		(0.400)
Single Child (Dummy)		-0.541		0.0116
		(0.555)		(0.366)
Collectivistic		0.456		-0.110
		(0.520)		(0.506)
Individualistic		0.257		0.868**
		(0.513)		(0.435)
Trustworthy (Belief)		-0.110		0.0408
		(0.112)		(0.108)
Public Order (Belief)		0.0689		-0.0807
		(0.0926)		(0.0808)
Period	-0.688	-0.378	0.151	0.273
	(0.708)	(0.705)	(0.544)	(0.520)
Period Squared	0.0234	0.0123	-0.0128	-0.0170
	(0.0279)	(0.0275)	(0.0223)	(0.0212)
Constant	-1.483	-3.893	-0.915	-3.338
	(4.524)	(5.051)	(3.215)	(3.493)
Observations	3534	3534	3985	3985

Notes: The dependent variable is the punishment points assigned to group member who contributed more or less than the punisher. The Perc. Paddy Field is the percentage of cultivated land devoted to rice paddy fields at the municipality level. and Anti-Social Punishment is a dummy variable which equals one if the punished subject contribute *more* then the punisher. Cluster Standard errors (PGG group level) are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

Table A.2: Random Effects Panel Regressions regarding the contribution in the PGGs.

	PGG without Punishment		PGG with Punishment	
	(1)	(2)	(3)	(4)
Rice	1.086*	1.021*	1.577**	1.437**
	(0.583)	(0.556)	(0.624)	(0.581)
GDP per cap. (Municipality)		-0.0316		0.0464
		(0.0596)		(0.0608)
Holistic Thinking		1.328*		0.311
		(0.766)		(0.734)
From Rural (Dummy)		-0.275		-0.410
		(0.410)		(0.412)
Relative Income		-0.564**		-0.295
		(0.279)		(0.237)
Risk Attitude		0.0726		0.159
		(0.117)		(0.106)
Male		2.029***		2.159***
		(0.462)		(0.369)
Natural Science (Dummy)		0.653*		1.313***
		(0.382)		(0.392)
Priming (Dummy)		-0.468		-0.158
		(0.541)		(0.567)
Single Child (Dummy)		-0.364		-0.0905
		(0.478)		(0.409)
Collectivistic		0.488		0.676
		(0.443)		(0.465)
Individualistic		-0.629		-0.431
		(0.427)		(0.388)
Trustworty (Belief)		0.0777		0.123
		(0.0897)		(0.104)
Public Order (Belef)		-0.0336		-0.0448
		(0.0922)		(0.0796)
Period	1.191***	1.191***	2.244***	2.244***
	(0.171)	(0.171)	(0.301)	(0.302)
Period sqaured	-0.144***	-0.144***	-0.0746***	-0.0746***
	(0.0173)	(0.0174)	(0.0115)	(0.0115)
Constant	7.911***	7.365***	-4.147**	-7.435**
	(0.419)	(2.544)	(1.772)	(3.057)
Observations	4112	4112	4112	4112

Notes: The dependent variable is contribution in the PGG without punishment (columns 1 and 2) and with punishment (columns 3 and 4). Rice is a dummy variable equals one if the subject comes from a Rice farming province. Cluster Standard errors (PGG group level) are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

Table A.3: Tobit Regressions regarding amount of punishment assigned to free-riders and cooperators.

	PGG without Punishment		PGG with Punishment	
	(1)	(2)	(3)	(4)
Rice	1.086*	1.021*	1.577**	1.437**
	(0.583)	(0.556)	(0.624)	(0.581)
GDP per cap. (Municipality)		-0.0316		0.0464
		(0.0596)		(0.0608)
Holistic Thinking		1.328*		0.311
		(0.766)		(0.734)
From Rural (Dummy)		-0.275		-0.410
		(0.410)		(0.412)
Relative Income		-0.564**		-0.295
		(0.279)		(0.237)
Risk Attitude		0.0726		0.159
		(0.117)		(0.106)
Male		2.029***		2.159***
		(0.462)		(0.369)
Natural Science (Dummy)		0.653*		1.313***
		(0.382)		(0.392)
Priming (Dummy)		-0.468		-0.158
		(0.541)		(0.567)
Single Child (Dummy)		-0.364		-0.0905
		(0.478)		(0.409)
Collectivistic		0.488		0.676
		(0.443)		(0.465)
Individualistic		-0.629		-0.431
		(0.427)		(0.388)
Trustworty (Belief)		0.0777		0.123
		(0.0897)		(0.104)
Public Order (Belef)		-0.0336		-0.0448
		(0.0922)		(0.0796)
Period	1.191***	1.191***	2.244***	2.244***
	(0.171)	(0.171)	(0.301)	(0.302)
Period sqaured	-0.144***	-0.144***	-0.0746***	-0.0746***
	(0.0173)	(0.0174)	(0.0115)	(0.0115)
Constant	7.911***	7.365***	-4.147**	-7.435**
	(0.419)	(2.544)	(1.772)	(3.057)
Observations	4112	4112	4112	4112

Notes: Tobit Regressions. The dependent variable is the punishment points assigned to group members who contributes less or more than the punisher. Rice is a dummy variable equals one if the subject comes from a Rice farming province. Anti-Social punishment is a dummy variable which equals 1 if the punished subject contributes more than the punisher. Cluster Standard errors (PGG group level) are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

Table A.4: Comparing the Probability of Punishment and the Intensity of Punishment Conditional on Punishing between subjects from rice provinces

	Probabiliy of Punishment			Intensity of Punishment		
	(1)	(2)	(3)	(4)	(5)	(6)
Zhejiang	0.0641	0.0747	0.0579	-0.158	-0.240	-0.791*
	(0.0662)	(0.0651)	(0.0806)	(0.442)	(0.429)	(0.426)
Anti-Social Punishment	-0.141***	-0.0895***	-0.0891***	-0.564	-0.0602	0.0512
	(0.0296)	(0.0336)	(0.0338)	(0.377)	(0.407)	(0.335)
Zhejiang x Anti-Social Punishment	-0.0268	-0.0282	-0.0274	0.874	0.980	0.869*
	(0.0391)	(0.0383)	(0.0384)	(0.593)	(0.596)	(0.481)
Punisher Contribution		0.00851**	0.00798*		0.0669	0.0333
		(0.00423)	(0.00429)		(0.0537)	(0.0452)
Punished Contribution		-0.00853***	-0.00858***		-0.0359	-0.0346
		(0.00294)	(0.00299)		(0.0231)	(0.0231)
Other Two Member Avg Contribution		-0.0167***	-0.0168***		-0.00128	-0.00167
		(0.00476)	(0.00488)		(0.0382)	(0.0363)
GDP per cap. (Municipality)			-0.00253			0.0471
			(0.00748)			(0.0394)
Holistic Thinking			-0.136			0.222
			(0.106)			(0.671)
From Rural (Dummy)			0.00435			0.345
			(0.0461)			(0.326)
Relative Income			0.0200			0.390*
			(0.0255)			(0.218)
Risk Attitude			-0.00121			0.0957
			(0.0146)			(0.116)
Male			0.0632			1.091***
			(0.0440)			(0.368)
Natural Science (Dummy)			0.0120			0.0827
			(0.0389)			(0.291)
Priming (Dummy)			0.0572			-0.317
			(0.0592)			(0.353)
Single Child (Dummy)			-0.0232			-0.385
			(0.0440)			(0.345)
Collectivistic			0.0266			-0.0895
			(0.0537)			(0.290)
Individualistic			0.0223			0.155
			(0.0507)			(0.289)
Trustworthy (Belief)			-0.0137			0.00754
			(0.0117)			(0.0894)
Public Order (Belief)			0.00817			0.0775
			(0.00926)			(0.0862)
Period	-0.0723	-0.0269	-0.0239	0.586	0.558	0.357
	(0.0646)	(0.0672)	(0.0675)	(0.527)	(0.519)	(0.488)
Period Squared	0.00283	0.00125	0.00114	-0.0249	-0.0239	-0.0150
	(0.00251)	(0.00259)	(0.00260)	(0.0211)	(0.0209)	(0.0195)
Constant	0.841**	0.719*	0.629	-0.692	-1.114	-2.092
	(0.407)	(0.407)	(0.450)	(3.033)	(3.026)	(3.600)
Observations	3534	3534	3534	1223	1223	1223

Notes: Columns 1-3 are Random Effects Linear Probability models and the dependent variable is the probability of punishing Free-rider and Cooperators. Columns 4-6 are tobit regressions and the dependent variable is the amount of punishment points assigned to other group member who contributed more or less than the punisher. Cluster Standard errors (PGG group level) are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

Table A.5: Comparing the Probability of Punishment and the Intensity of Punishment Conditional on Punishing between subjects from non-rice provinces

	Probabiliy of Punishment			Intensity of Punishment		
	(1)	(2)	(3)	(4)	(5)	(6)
Shandong	-0.0166	-0.0344	-0.0212	0.322	0.345	0.563**
	(0.0491)	(0.0506)	(0.0566)	(0.341)	(0.304)	(0.223)
Anti-Social Punishment	-0.147***	-0.0716**	-0.0716**	-0.247	0.288	0.289*
	(0.0207)	(0.0310)	(0.0310)	(0.201)	(0.205)	(0.174)
Shandong x Anti-Social Punishment	0.0156	0.00369	0.00287	0.201	0.265	0.141
	(0.0290)	(0.0310)	(0.0309)	(0.304)	(0.356)	(0.298)
Punisher Contribution		0.0129***	0.0132***		0.0707*	0.0492*
		(0.00449)	(0.00455)		(0.0426)	(0.0270)
Punished Contribution		-0.0121***	-0.0117***		-0.0513**	-0.0433**
		(0.00269)	(0.00266)		(0.0245)	(0.0171)
Other Two Member Avg Contribution		-0.0215***	-0.0209***		-0.0618*	-0.0500**
		(0.00361)	(0.00354)		(0.0357)	(0.0236)
GDP per cap. (Municipality)			-0.00290			-0.0232
			(0.00629)			(0.0346)
Holistic Thinking			-0.165*			-0.0594
			(0.0865)			(0.286)
From Rural (Dummy)			-0.00259			0.516**
			(0.0474)			(0.218)
Relative Income			-0.0102			-0.129
			(0.0344)			(0.143)
Risk Attitude			0.0141			0.175**
			(0.00995)			(0.0861)
Male			-0.0301			0.260
			(0.0325)			(0.204)
Natural Science (Dummy)			-0.0232			0.316
			(0.0383)			(0.198)
Priming (Dummy)			0.0973**			0.0667
			(0.0457)			(0.234)
Single Child (Dummy)			-0.0152			0.313
			(0.0400)			(0.277)
Collectivistic			-0.0262			0.170
			(0.0452)			(0.341)
Individualistic			0.0801*			0.469**
			(0.0479)			(0.211)
Trustworthy (Belief)			0.00188			-0.000822
			(0.0114)			(0.0566)
Public Order (Belief)			-0.00696			-0.0145
			(0.00838)			(0.0507)
Period	0.0228	0.0660	0.0644	-0.335	-0.188	-0.343
	(0.0644)	(0.0627)	(0.0626)	(0.230)	(0.242)	(0.250)
Period Squared	-0.00158	-0.00312	-0.00306	0.0110	0.00593	0.0118
	(0.00259)	(0.00253)	(0.00253)	(0.00957)	(0.0101)	(0.0100)
Constant	0.329	0.254	0.189	4.274***	3.441**	1.210
	(0.385)	(0.392)	(0.428)	(1.422)	(1.467)	(1.449)
Observations	3985	3985	3985	1156	1156	1156

Notes: Columns 1-3 are Random Effects Linear Probability models and the dependent variable is the probability of punishing Free-rider and Cooperators. Columns 4-6 are tobit regressions and the dependent variable is the amount of punishment points assigned to other group member who contributed more or less than the punisher. Cluster Standard errors (PGG group level) are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

A.3 Experimental Instructions

Thank you for participating.

Please note that communication with other participants is prohibited during the study. If you have a question once the study has begun, please raise your hand and an assistant will come to your desk to answer it. Violation of this rule can lead to immediate exclusion from the study and from all payments.

Today we will do 5 studies. The instructions for the first study are attached. Once a study is completed, you will receive instructions for the next study.

During the study we will not speak in terms of GBP, but in points. Your entire earnings will be calculated in points. At the end of the study the total amount of points you have earned will be converted to RMB at the following rate:

1 point = 0.4 GBP

At the end of today's study, one out of 5 study will be randomly selected for payment. After you completed all the studies, a card will be drawn from a bag, containing cards numbered from 1 to 5. The number on the card determines which study is for payment.

You will receive GBP 4 as a show-up fee for participating. Therefore, your total earning is:

Total Earning = Show-up fee + money you earned in the randomly chosen study

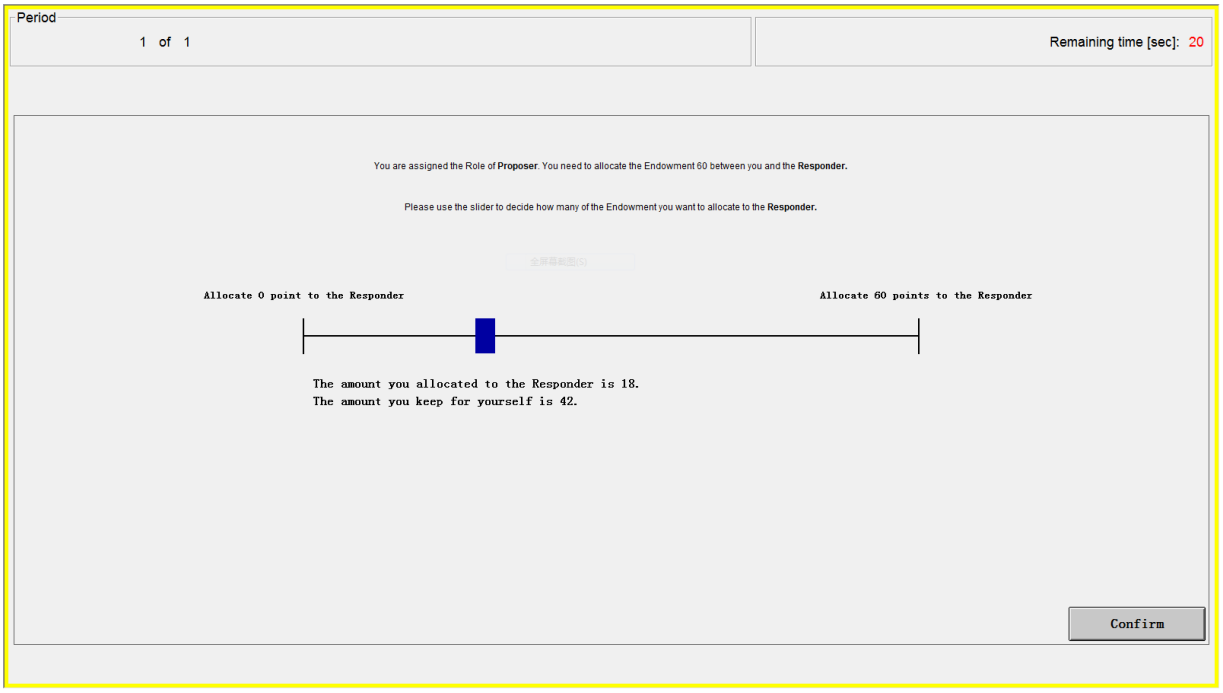
Please read the instructions carefully, because your earnings in each study depends on how well you understand the instructions.

Instructions for the First Study

In this study, first you will be assigned a role. You will be either a Proposer or a Responder. If you are a Proposer, you will be randomly and anonymously paired with a Responder. If you are a Responder, you will be randomly and anonymously paired with a Proposer. This way, half of the people in the room will be Proposers and half of them will be Responders.

DECISION OF PROPOSER

The Proposer's role is to allocate **a total of 60 points** between the Proposer and Responder. The input screen for the Proposer is presented below:



The Proposer needs to use the slider to allocate points between him or her and the Responder. The more points the Proposer allocates to the Responder the less points he or she keeps. The amount of points allocated to the Responder as well as the points remaining for the Proposer are both shown on the screen.

DECISION OF RESPONDER

In the current study the responder can only accept the allocation made by the Proposer. In other words, the allocation made by the Proposer is implemented regardless of whether the Responder agrees or disagrees.

EARNINGS

The Proposer and the Responder receive the amount according to the allocation made by the Proposer.

Control questions

1. Suppose the Proposer allocated 20 points to the responder.

What are the earnings for the Proposer?.....

What are the earnings for the Responder?.....

What can the Responder do if he/she is not satisfied with the allocation? ……..

……

2. Suppose the Proposer allocated 40 points to the responder.

What is the earnings for the Proposer?.....

What is the earnings for the Responder?.....

What can the Responder do if he/she is not satisfied with the allocation? ……..

……

Instructions for the Second Study

This study is very similar to the previous one.

Your role in this study remains the same as in the previous study. If you were a Proposer, you will also be a Proposer in this one. If you were a Responder, you will also be a Responder in this one.

Again . If you are a Proposer, you will be randomly and anonymously paired with a Responder. If you are a Responder, you will be randomly and anonymously paired with a Proposer. This way, half of the people in the room will be Proposers and half of them will be Responders. Your pair in this study **need not be** the same as in the previous study.

DECISION OF PROPOSER

The decision of the Proposer is exactly the same as in the previous study. The Proposer needs to allocate a total of 60 points between the Proposer and the Responder. In this study, the Responder can accept or reject the offer.

NEW IN STUDY 2: DECISION OF RESPONDER

Responders need to enter the **minimum acceptance amount** while the Proposers are making their decisions. The minimum acceptance amount is a number such that if the Proposer allocates a number **less** than the minimum acceptance amount, the allocation will be **automatically rejected**. On the other hand, if the Proposer allocates a number **more or equal** to the minimum acceptance amount, the allocation will be **automatically accepted**. For example, if a Responder stated 20 as the minimum acceptance amount and the Proposer allocates 19 or less to the Responder, then the allocation is automatically rejected. If the Proposer allocates 20 or more points to the Responder, then the allocation is automatically accepted. Important, Responders and Proposers are making decisions simultaneously. Therefore, Proposers will **NOT** know Responders' minimum acceptance amount while making the allocation. Similarly, Responders will **NOT** know Proposers' allocation while entering minimum acceptance amount.

While Proposers are making decisions, Responders need to enter a number between 0 and 60. This number is called the "**Minimum Acceptance Amount.**" If the points that Proposer allocated to Responder are **less** than this "Minimum Acceptance Amount", Proposer's allocation will be **automatically rejected**. On the other hand, if the points that Proposer allocated to Responder are **more or equal** to this "Minimum Acceptance Amount", Proposer's allocation will be **automatically accepted**. For example, if a Responder stated 20 as the minimum acceptance amount and the Proposer allocates 19 or less to the Responder, then

the allocation is automatically rejected. If the Proposer allocates 20 or more points to the Responder, then the allocation is automatically accepted. Important, Responders and Proposers are making decisions simultaneously. Therefore, Proposers will **NOT** know Responders' minimum acceptance amount while making the allocation. Similarly, Responders will **NOT** know Proposers' allocation while entering minimum acceptance amount.

The input screen for the Responder is presented below.

EARNINGS

If the allocation made by Proposer is **accepted**, both receives the points allocated to them.

If the allocation made by Proposer is **rejected**, both receive **zero points**.

Please answer the questions in the next page. They serve as a test for you understanding of the task.

Control questions

1. Suppose the Proposer allocated 20 points to the Responder.

If Responder enter the minimal acceptance amount 15, what is the earnings for the Proposer?.....

What are the earnings for the Responder?.....

If Responder enter the minimal acceptance amount 45, what is the earnings for the Proposer?.....

What are the earnings for the Responder?.....

2. Suppose the Proposer allocated 40 points to the responder.

If Responder enter the minimal acceptance amount 15, what is the earnings for the Proposer?.....

What are the earnings for the Responder?.....

If Responder enter the minimal acceptance amount 45, what is the earnings for the Proposer?.....

What are the earnings for the Responder?.....

Instructions for the Third Study

In this study, participants are randomly divided into groups of two. You will therefore be in a group with another participant.

DECISIONS

You and the other participant in your group need to pick one out of two possible choices simultaneously. The choices are labelled @ and #. When you make your choice you will not know what the other participant will choose. The other participant will not know your choice either. In other words, no participant will know what action the other player chose when making a decision.

EARNINGS

The following table shows earning for all possible combination of choices made by you and the other participant in your group.

		Other's Choice	
		@	#
Your Choice	@	(**30**, 30)	(**10**, 22)
	#	(**22**, 10)	(**22**, 22)

Note that, the numbers that are **Bolded** in each cell are **earnings for you**. The other number in each cell indicates the earning for the other participant.

For example, suppose your choice is "@" and the other's choice is "#", then the earning are (**10**, 22). Therefore, you earn **10** points and the other participant earns 22 points. If you choose "#" and the other's choice is "#", then the earning are (**22**, 22). Therefore, you earn **22** points and the other participant earns 22 points too.

Keep in mind: You and the other participant make your choices simultaneously without knowing what the other participant chooses.

Control questions:

1. Suppose you choose @ and the other participant choose @.

What is the earning for you?.....

What is the earning for the other participant?.....

2. Will you know what the other participant chose when you chose? ….

Will the other participant know what you chose once he or she choses? …

3. Suppose you earn 22 and the other participant earn 10.

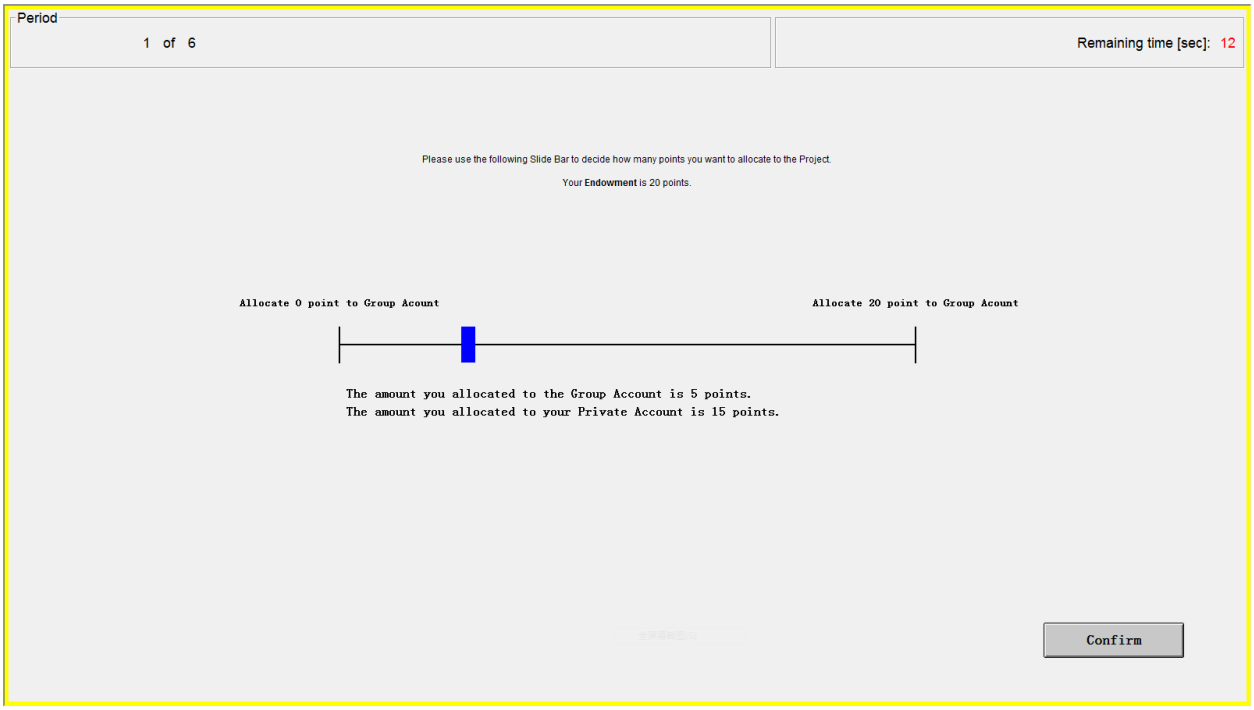
What was your choice?.....

What was the other participant's choice?.....

Instructions for the Fourth Study

In this study, participants are randomly divided into groups of four. You will therefore be in a group with 3 other participants. You will remain in the same group for the duration of this study. There will be a total of 6 periods, each participant will face the same decision in each period.

At the beginning of each period, each participant receives 20 points. We call this your endowment. In each period you will be asked to decide how many points of your endowment you want to allocate to a **Group Account.** You may allocate any integer number of points between 0 and 20. The remainder of your endowment will be automatically allocated to your **Individual Account**. The input screen is presented below:



You can use the slide bar to decide how many points of your endowment you want to allocate to the **Group Account**. The amount allocated to your **Private Account** is also shown on the screen.

EARNINGS

After all the participants have made their decisions, your earnings for the period are calculated. Your earnings consist of two parts:

(1) Your earnings from the **Individual Account**.

(2) Your earnings from the **Group Account.**

Your earnings from the Individual Account equal the points that you keep for yourself, and are thus independent of others' decisions. For every point you keep for yourself in your Individual Account, you earn 1 point.

Your earnings from the Group Account depend on the total number of points allocated to the Group Account by the 4 group members (including yourself). This total amount is multiplied by 1.6 and then distributed equally amongst the four group members – each member receives a quarter of it (25%). In other words, each point that you allocate to the Group Account turns into 1.6 points, which are distributed equally to four members i.e. 0.4 points each.

So, for each point that you or any of your group members allocate to the Group Account, you and the other three group members receive 0.4 points each.

In summary, your earnings in each period are calculated as follows:

Your earnings =

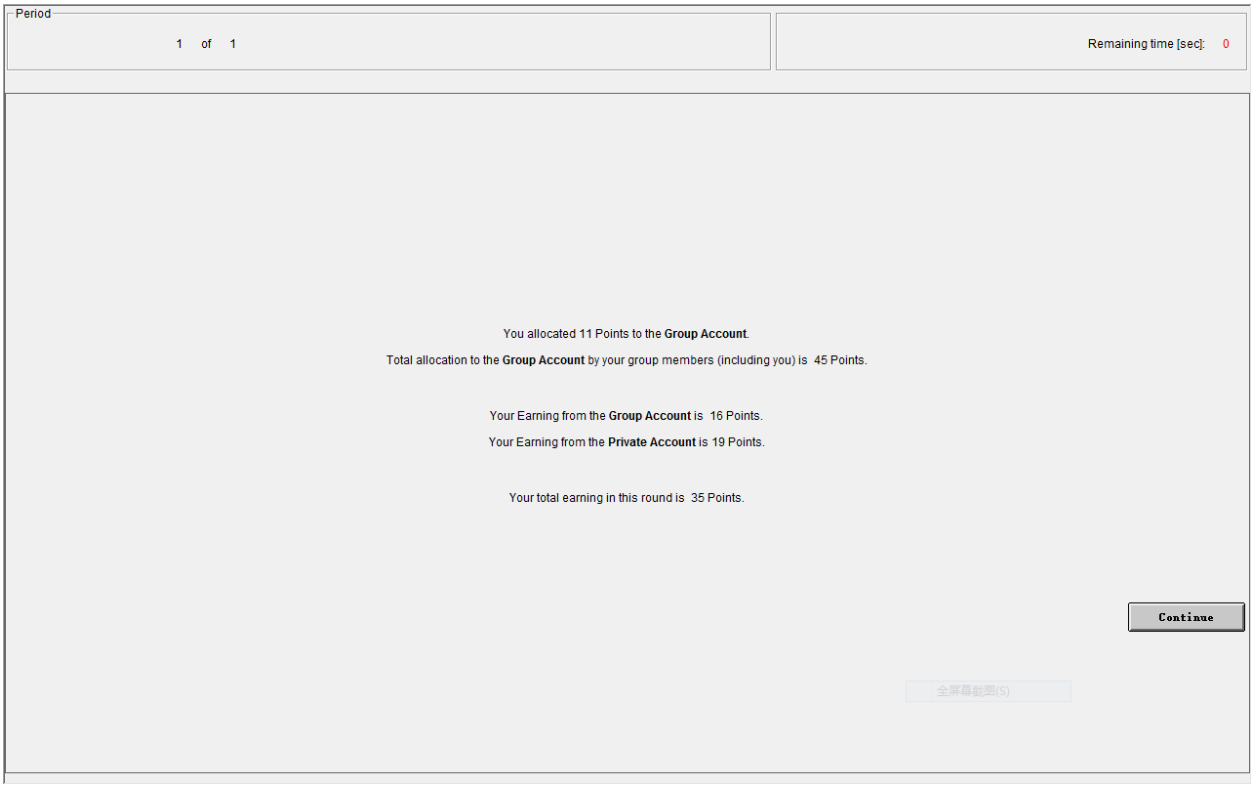
Earnings from Individual Account + Earnings from the Group Account =

20 - (Your allocation to the Group account) + 0.4 x (Total points allocated to Group Account by all group members)

Example: Suppose in one period that you allocated 8 points to the Group Account and that the other three members of your group allocated a total of 22 points. This makes a total of 30 points in the Group Account. In this case each member of the group receives earnings from the Group Account of 0.4×30 = 12 points. In addition, you also receive 12 points from your Individual Account. Therefore, your earning in this period is: (20 – 8) + 0.4 x 30 = 24 points.

RESULTS SCREEN

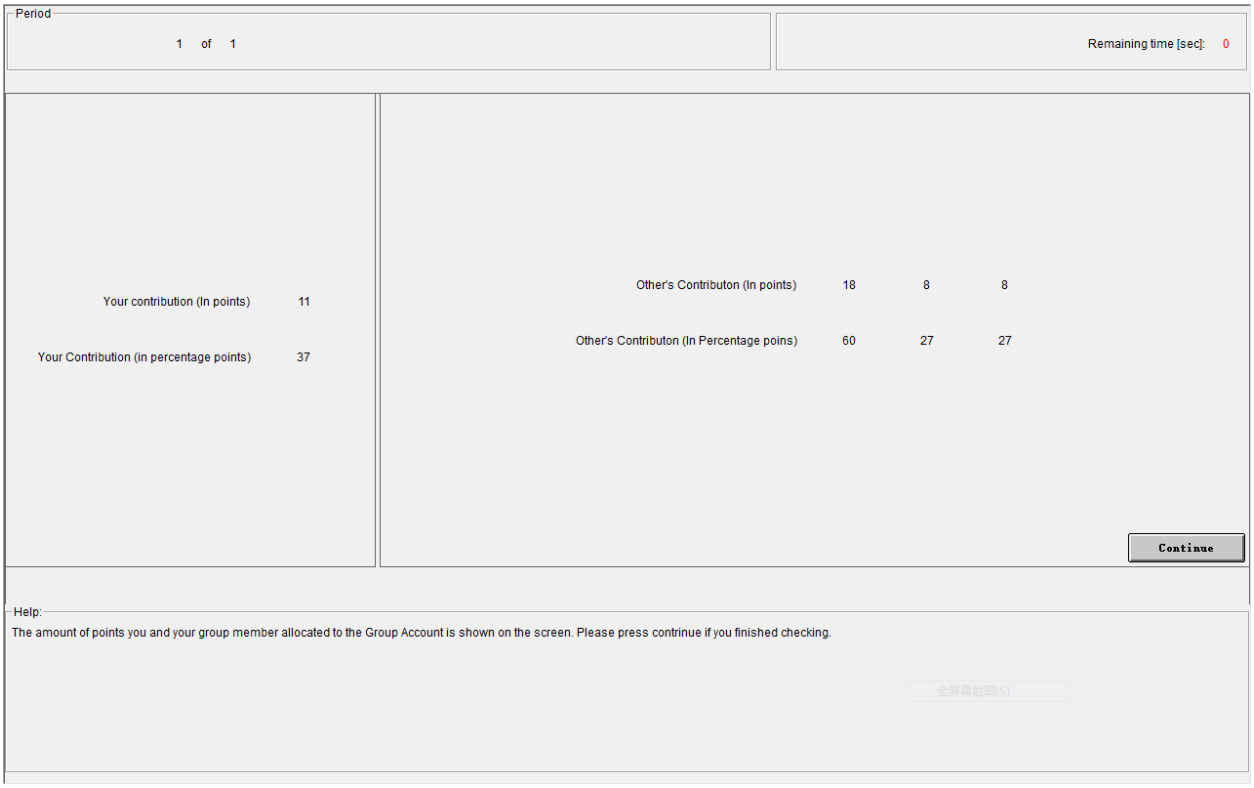
After all your group members have made their decision, your allocation and the sum of all allocations in your group are reported on the Result Screen as shown below. To aid you in your calculation, your earnings from your individual account and your earnings from the group account are both presented on the screen.



Please press the Continue button after you have read all the information.

INFORMATION SCREEN

Next the information screen appears, which reveals the contributions of the other group members.



This screen shows how many points each group member allocated to the Group Account. Your allocation is displayed in the first column, while the allocations made by the other group members are shown in the remaining three columns. Please note that **the order in which other group member's allocations are displayed changes randomly in every period.** The allocation in the second column, for example, generally represents a different group member each time. The same holds true for the allocations in the other columns. **That way you are informed about the contributions but not about the identities of the other group members.**

A new period will start shortly after pressing the Continue button. You will again receive 20 points as endowment and you will be asked again to decide how many points of your endowment you want to allocate to a Group Account.

If this study is randomly chosen for payment, we will randomly pick 1 period out of the 6 periods and your payments will be calculated by your decisions in that period

Please answer the questions in the next page. They serve as a test for your understanding of the task.

Control questions

1. Each group member has an endowment of 20 points. Suppose nobody (including you) contributes any points to the Group Account. What is:

Your earnings from the Group Account?.....

Your earnings from the Individual Account?.....

Your total earnings?.....

Other group members earnings from the Group Account?.....

Other group members earnings from the Individual Account?.....

Other group members total earnings?.....

2. Each group member has an endowment of 20 points. Suppose you contribute 8 points to the Group Account. All other group members each contribute 12 points to the Group Account. What are:

Your earnings from the Group Account?.....

Your earnings from the Individual Account?.....

Your total earnings?.....

Other group members earnings from the Group Account?.....

Other group members earnings from the Individual Account?.....

Other group members total earnings?.....

3. Each group member has an endowment of 20 points. Suppose the other three group members contribute **a total of 30** points to the Group Account.

a) If you contribute 5 points to the Group Account.

Your earnings from the Group Account?.....

Your earnings from the Individual Account?.....

Your total earnings?.....

Other group members earnings from the Group Account?.....

b) What are your earning if you contribute 15 points to the Group Account?

Your earnings from the Group Account?.....

Your earnings from the Individual Account?.....

Your total earnings?.....

Instructions for the Fifth (Last) Study

This study is similar to the previous study. First you will be randomly divided into a new group of four. The **new** group composition will **not** change throughout this study.

Each participant receives a lump sum payment of **10 Points** at the beginning of this study. This one-off payment can be used to pay for eventual losses during this study**. However, you can always evade losses with certainty through your own decisions**.

This study consists of 10 periods and there are **2 stages in each period**. The first stage is identical to the previous study. At the beginning of each period each participant receives 20 points as his or her endowment. You need to decide how many points of your endowment you want to allocate to a **Group Account** (and hence the remainder of your endowment will be automatically allocated to your **Individual Account**). Your earnings from the first stage will be calculated exactly in the same way as in the previous part.

Your earnings from the **First Stage** =

Earnings from Individual Account + Earnings from the Group Account =

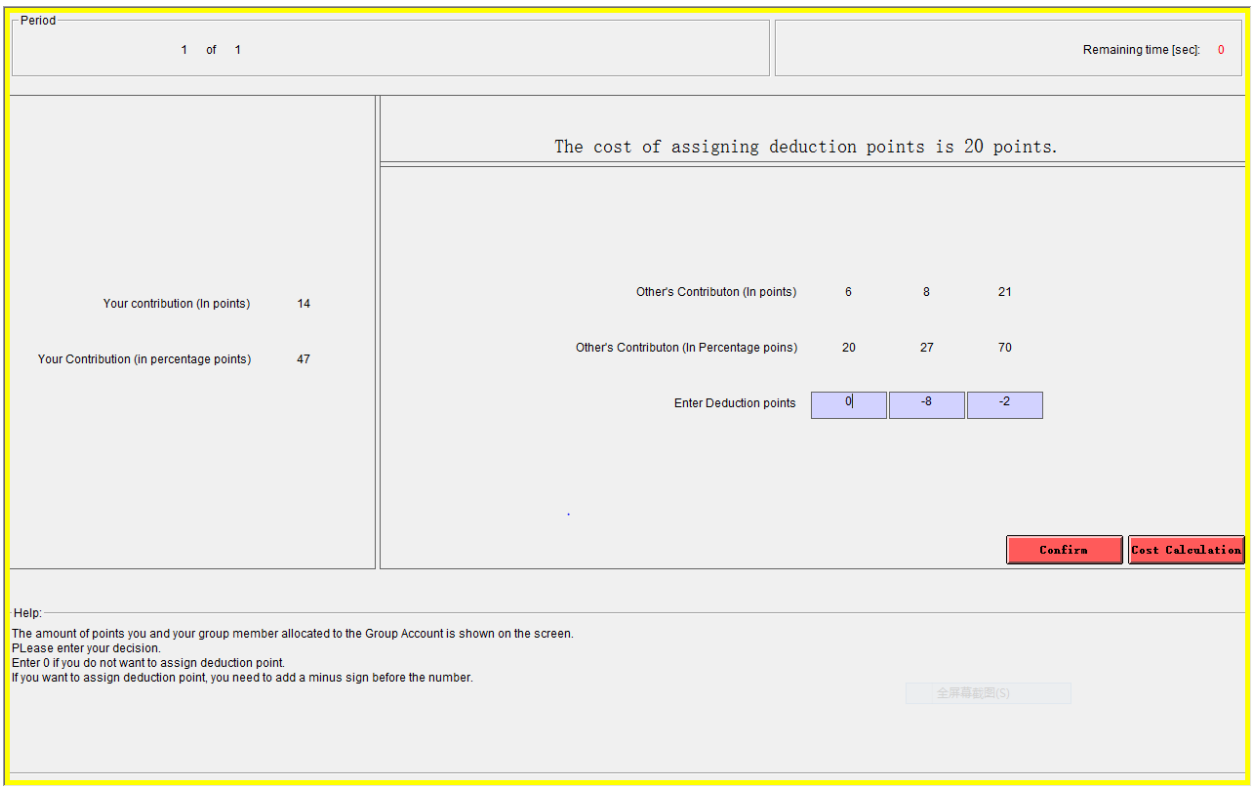
20 - (Your allocation to the Group account) + 0.4 x (Total points allocated to Group Account by all group members)

THE SECOND STAGE

There will be a new **second stage** introduced after all participants have made their decisions in the first stage.

At the second stage you can observe how many points each group member allocated to the Group Account. In addition, in this stage you can **decrease** the earning of each group member by assigning **deduction tokens** to him/her. If you do not want to decrease the other's earning, you simply do not assign any deduction tokens to him/her. Note that other group members can also decrease your earnings if they wish to do so.

The input screen for the second stage is presented below:



The screen shows how many points each group member allocated to the Group Account at the first stage. Your allocation is displayed in the first column, while the allocations made by the others are shown in the remaining three columns. Please note that **the order in which allocations are displayed changes randomly in every period**. The allocation in the second column, for example, generally represents a different group member each time. The same holds true for the other columns. This way you are informed about the contributions but not about the identities of the other group members.

You now have to decide whether, and if so how many, deduction tokens to assign to each of the other three group members. If you do not wish to change the income of a specific group member then you must enter 0. If you want to distribute deduction tokens, you must put a negative sign in front of the number (without spaces between them).

You can assign between 0 and 10 deduction tokens to each group member. However, each deduction token **costs you 1 point**. Therefore, the larger the amount of deduction tokens that you assign to other group members, the larger your costs. The total cost of assigning deduction tokens is calculated as follows:

Total cost of assigning deduction tokens = Sum of assigned deduction tokens x 1

You can move from one input field to the other using the mouse.

Example: If you assign 2 deduction tokens to one member (enter -2), assign 8 deduction tokens to another member (enter -8), and you assign 0 deduction token to the last group member (enter 0), the sum of assigned deduction tokens is 2 + 9 + 0 = 11 and the total cost is 11 x 1 = 11 points.

Each deduction token assigned to a participant reduces his/her earnings by 3 points. A participant's total received deduction tokens equal the sum of deduction tokens other group members assigned to him/her. Consequently, the amount of earnings decreased by the received deduction tokens is calculated as follows:

Total amount of earnings decreased by received deduction tokens = Sum of received deduction tokens x 3

Important: By receiving deduction tokens, each participant's earning can only be reduced to **ZERO**.

Example: If a participant received 2 deduction token from one group member, 9 deduction tokens from another group member, and 0 deduction token from the last group member, then the participant received a total of 2 + 9 + 0 = 11 deduction tokens. Consequently, his/her earnings will be decreased by 11 * 3 = 33 points. If this participant earned 40 points in the **First Stage**, then his/her earnings will be 40 – 33 = 7 points**. If this participant earned less than 33 in the First Stage, his/her earning will only be reduced to 0 point.** It is possible that one can earn a negative amount: if your earnings were reduced to ZERO by receiving deduction tokens **and you distributed 5 deduction tokens to others**, your final earnings will be 0 – 5 = -5 points. **However, you can always evade losses with certainty through your own decisions**.

EARNINGS

After all participants have made their decisions in the second stage, your earnings for the period are calculated.

The earnings from the First Stage are the same as in the previous part. These are the earnings from your Individual Account and the earnings from the Group Account.

The earnings from the **Second Stage** depend on the total deduction tokens you assigned to other group members as well as the total deduction tokens you received from other group members.

In sum, your earnings in each period are calculated as follows:

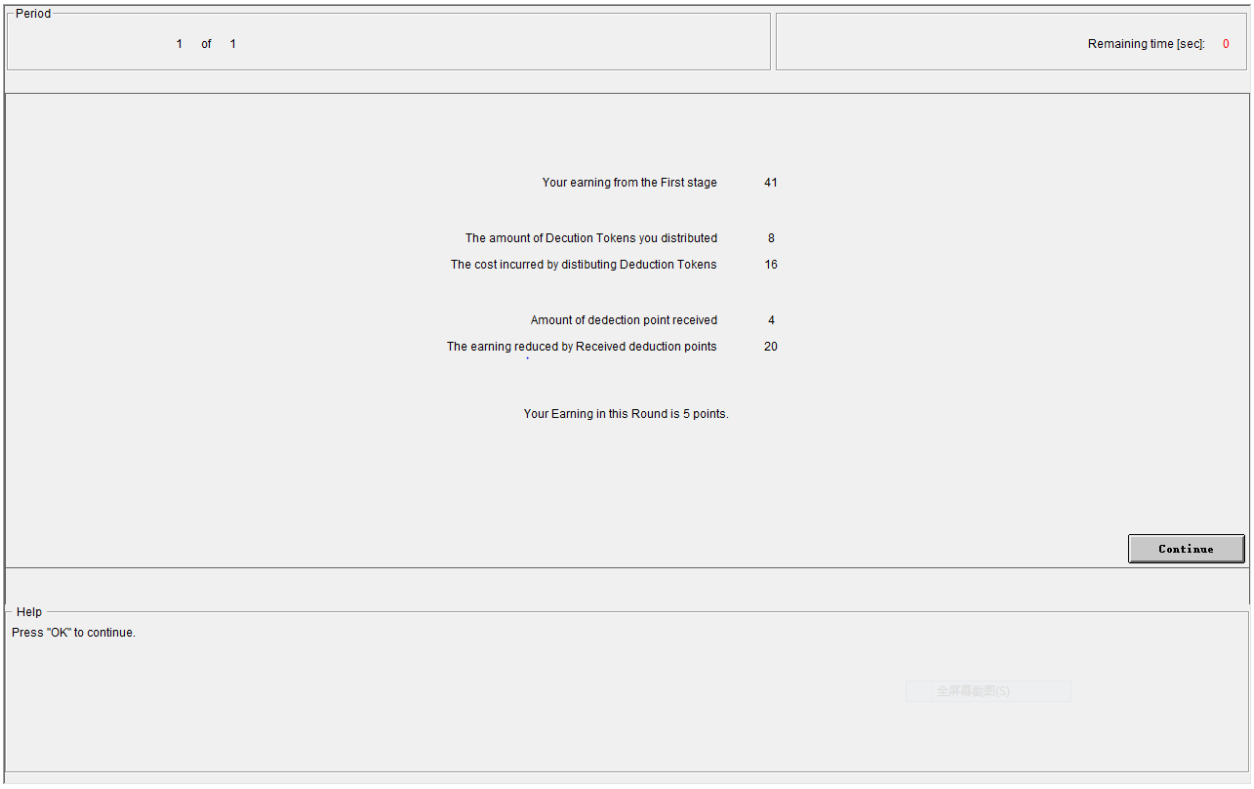
Your earnings at the end of the second stage = income per period

= **Earnings in the First stage**
- (Sum of deduction tokens received from other participants x 3)
- (Sum of deduction tokens assigned to other participants)

Please remember that your earnings at the end of the second stage can be negative, if the cost of your points used to distribute deduction tokens exceeds your (possibly reduced) income from the first stage. You can however avoid such losses with certainty through your own decisions!

RESULTS SCREEN

At the end of the second stage, your allocation and the sum of all allocations in your group are reported on the outcome screen as shown below. The sum of deduction tokens you assigned to others as well as the sum of deduction tokens you received are also presented on the screen.



Please press the Continue button after you have read all the information. A new period will start shortly.

If this study is randomly chosen for payment, we will randomly pick 1 period out of the 10 periods and your payments will be calculated by your decisions in that period

Please answer the questions in the next page. They serve as a test for you understanding of the task.

Control questions

1. Suppose at the second stage you assign the following deduction tokens to your three other group members: -9, -5, and 0. What is the total cost of your assigned deduction tokens?.....

2. What is your cost if you assign a total of 0 points?.....

3. Suppose you earn 10 points in the First stage. By how many points will your income from the first stage be reduced if you receive a total of 1 deduction tokens from the other group members?.....

4. Suppose you earn 20 points in the First stage. By how many points will your income from the first stage be reduced if you receive a total of 5 deduction tokens from the other group members?.....

5. Suppose you earn 30 points in the First stage. If you received 1 deduction token and assigned a total of 5 deduction tokens. What are your final earnings?.....

6. Suppose you earn 20 points in the First stage. If you received 7 deduction tokens and assigned a total of 5 deduction tokens. What are your final earnings?.....

7. Suppose you earn 10 points in the First stage. If you received 2 deduction tokens and assigned a total of 8 deduction tokens. What are your final earnings?.....

B Appendix to Chapter 3

General instructions for participants

Thank you for participating in this study.

Please note that it is prohibited to communicate with other participants during the study. If you have a question once the study has begun, please raise your hand and an assistant will come to your desk to answer it. Violation of this rule leads to immediate exclusion from the study and from all payments.

You will never learn the identity of the other participants, neither before nor after the study; and none of the other participants will learn anything about your identity.

You will receive 200 Rand for participating in this session. The study consists of 4 parts ("Part I", "Part II", etc.), each involving a simple decision task, note that you cannot earn money for yourself in all 4 parts.

Your payment = 200 Rand

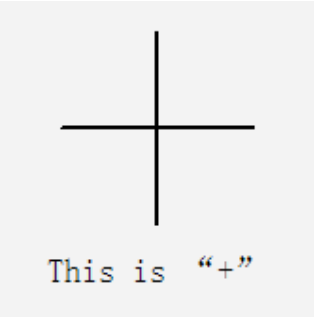
In what follows we describe the procedure for Part I in which no additional earnings for yourself are possible.

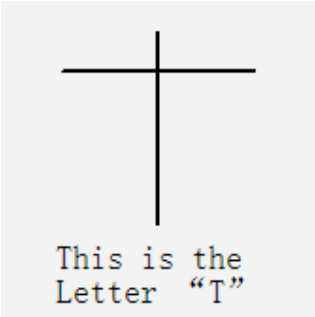
Part I

This part is similar to a "Run for charity" event. Instead of running you can solve some tasks and the more tasks you solve the more we donate to a charity of your choice. At the end of the study the supervisor will make an online donation to your chosen charity. Thus, the amount of money that you raise for the charity depends on your success in solving the tasks. Note, that you cannot earn any money for yourself in this part.

Note, earlier studies have shown that people can solve 19 tasks with little effort, and can solve 26 or more tasks with reasonable effort.

Below is the detailed description of the task:
The tasks that you are asked to solve are called "Find the T's". When the task starts, a page full of symbols will appear on your computer screen. There are just two different symbols on the screen, one is "+", the other is more similar to the letter "T".





In each task you need to find all the "T"s. Each task contains 20 "T"s in total. To select a "T" you simply left click with the mouse on the T. The "T"s that you selected will be highlighted in red. After you have found all of the "Ts" a new screen with new symbols will appear on your screen.

You have a total of 20 minutes to solve tasks. Note that this does not mean you need to solve tasks all the time. You can stop at any time by pressing the "STOP" button that is present on each screen. After pressing the stop button, you are given a choice to either continue solving the task (but there would be NO more money generated for the charity) or to stop this task completely.

The more tasks you solve the more is transferred to your chosen charity. The amount per solved task that is transferred to the charity decreases over time. Here is the list of donations that can be achieved:

Task 1:	3250 cents (=32.50 Rand)
Task 2:	1562 cents
Task 3:	1020 cents
Task 4:	729 cents
Task 5:	562 cents
Task 6:	458 cents
Task 7:	375 cents
Task 8:	312 cents
Task 9:	271 cents
Task 10:	229 cents
Task 11:	208 cents
Task 12:	167 cents
Task 13:	146 cents
Task 14:	146 cents
Task 15:	125 cents
Task 16:	104 cents
Task 17:	104 cents
Task 18:	83 cents
Task 19:	83 cents
Task 20 onwards:	63 cents

Your total donation will depend on the **Total** number of tasks solved. To calculate the total donation you simply need to add up the achieved donations of all solved tasks.

For example if you solved task 1 – 3, your donation will be:
3250 + 1562 + 1020 = 5832 cents = 58.32 Rand

If the tasks ends and you have not yet completed a given task the donation is calculated proportionally. This means that if you for example found 10 of the 20 Ts in task 6 your donation from task 6 will be half of the normal donation that is 229 cents instead of 458 cents. This means that no effort is lost when you stop the task or the time is up.

After 20 minutes or after pressing the "STOP" button, you can choose to which charity you would like the money to be sent that you raised.

Important rule:

It is important that you only click on the "Ts" and not on the "+" symbol. The reason is that if you click 4 times on a "+" symbol all Ts found so far in a task are deselected. That is you need to restart this task again. It does however not affect the previously solved tasks. The number of mistakes in a task does not carry on to the next task. That is, once you complete a task, the number of mistakes is reset to zero. The total number of mistakes is shown at the bottom of the screen.

Solving a task means that you need to find all the "T"s on that task. For example, if you are on task 26 and find less than 20 "T"s, this means that you did NOT solve 26 tasks but only solved 25 tasks.

Please choose one of the following charities. A brief description of each charity can be found on the next pages.

- O Lifeline
- O The South African Guide Dogs Association
- O Project Literacy
- O AIDS Foundation SA
- O Save the Rhino
- O St Luke's Hospice
- O Nazareth House Cape Town
- O World Wildlife Fund (WWF) South Africa
- O Students' Health and Welfare Centers Organization
- O The Sunflower Fund
- O Rape Crisis
- O Food & Trees for Africa (FTFA)
- O Southern African Foundation for the Conservation of Coastal Birds
- O Ihata Women's Shelter

Please feel free to ask questions at any point if you feel you need clarification. Please do so by raising your hand.

Control Questions

1. Suppose you finished Task 1 and you found 10 Ts on Task 2 and then you press the "Stop" button.

   How many Ts you did you find in total? ……

   How much money will you donate to the chosen charity? ……..

2. Suppose you have already found 6 Ts on a Task. Now you make several mistakes. What happens if you make …

   … 3 mistakes? ……..

   … 4 mistakes? ……..

3. Suppose you made 3 mistakes on Task 1. Now you are working on Task 2. How many mistakes you are allowed to make on Task 2?

   ……

4. It is possible to solve the tasks without generating any money for the

   charity? ……

   If yes, how can you do this? ……

   ……..……

Part II

In the second part, everything is the same as in part I except that we will publicly announce the seat number of those who solved 26 tasks or more.

Please note: because we are interested in why people donate in general, we will communicate the total amount raised by the group in this room at the end. We will ask those of you who solved 26 or more tasks to stand up. We will briefly acknowledge your efforts.

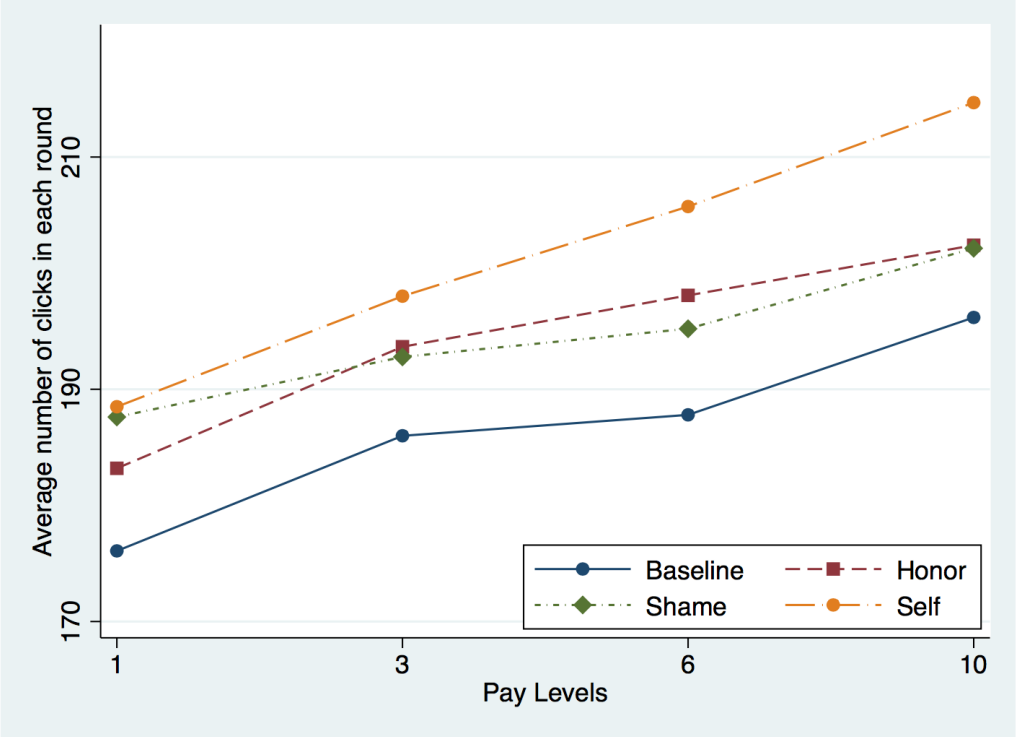
Important: We will <u>not</u> identify people with names, only with their seat number, which you find on the cubicle wall, and the amount donated. We will also <u>not</u> mention the charity people have chosen.

Remember: Solving a task means that you need to find all the "T"s on that task. For example, if you are on task 26 and find less than 20 "T"s, this means that you did **NOT** solve 26 tasks but only solved 25 tasks.

A Appendix to Chapter 4

A.1 Response to Different Pay Levels

Figure A.1: How Subjects Respond to Different Pay Levels



Notes: This figure illustrates that subjects on average put more effort when the pay level is higher. Importantly, subject's response to the different pay level is similar between the self condition and the donation condition. This implies that subject's behavior is rational when they cannot earn money for themselves.

A.2 Experimental Instructions

Thank you for participating in this study.

Please note that communication with other participants is prohibited during the study. If you have a question once the study has begun, please raise your hand and an assistant will come to your desk to answer it. Violation of this rule can lead to immediate exclusion from the study and from all payments.

Today, we will do two studies. The instructions for the first study are attached. The instructions for the second study will be provided upon completion of the first study. Payment of any earnings from either study will be made at the end of the second study, before you leave.

Instructions for the First Study

You will receive £8 as a show-up fee for participating in the first study.

The first study consists of two parts ("Part I" and "Part II"). Each part involves one simple task in which you can generate money that the experimenter will donate to a charity of your choice.

At the end, all participants will gather at the exit, and, after having signed a receipt, will receive their total payment (from both studies) in an envelope.

Furthermore, we will add your name to a list of all participants of this study. This list will be publicly viewable in the corridor outside of this room.

Please select one of the charities below. All donations generated by you will be given to your chosen charity. A brief description of each charity can be found on the next pages.

- Cancer Research UK
- Oxfam
- Greenpeace International
- British Red Cross
- British Heart Foundation
- ChildLine
- The Guide Dogs for the Blind Assoc
- World Wildlife Fund (WWF) UK
- RHUL: Scholarships & Bursaries
- Royal Soc. for the Protection of Birds

In the following we speak of points that you can donate to the charity. The points will be translated into money for the charity according to the following exchange rate:

100 points = 2 pence

In what follows we describe the procedure for Part I.

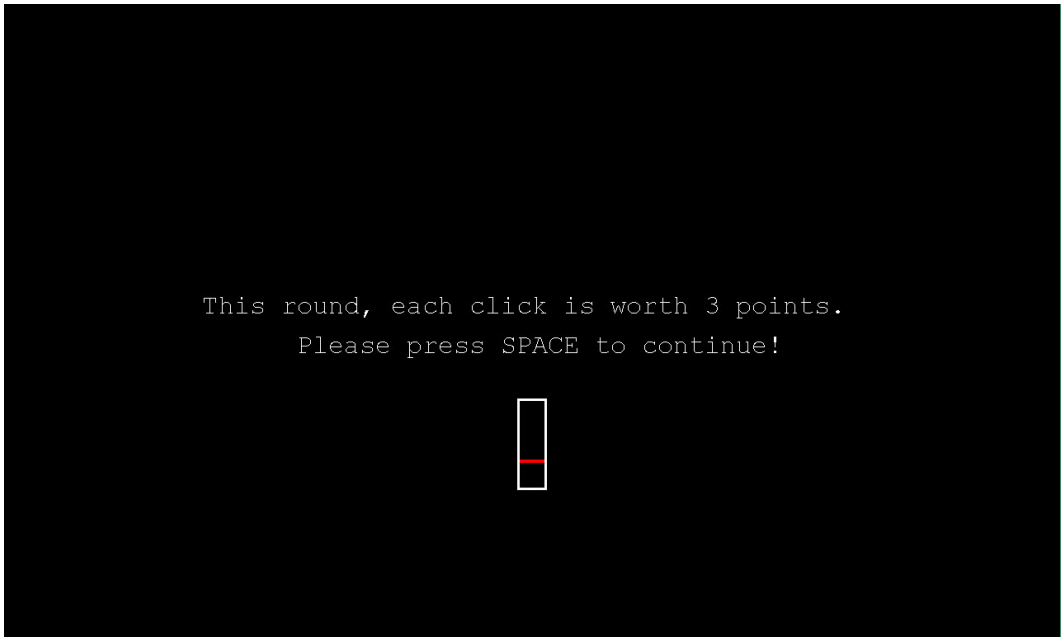
Part I: Click for Charity

This part is similar to a "Run for Charity" event, but instead of running, you click (with the LEFT BUTTON of your computer mouse). The more clicks you make, the more we will donate to a charity of your choice. We call this "Click for Charity."

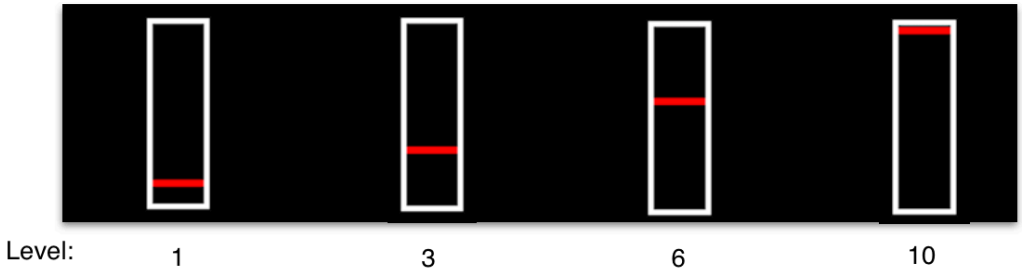
There will be 12 rounds, each following the same pattern.

Each round begins with a screen showing a "+" symbol. It indicates that the task will begin soon. No action is required.

After a short time, a new screen will appear. This screen will show how many points each click is worth in this round.



The amount of money you can generate for the charity of your choice is not the same for all rounds. In some rounds, one click is worth 1 point, in others 10 points.



There are a total of 4 different levels: 1, 3, 6, or 10 points per click. Level 1 means that in the current round, each click is worth 1 point. Similarly, level 3 means that in the current round, each click is worth 3 points.

When you are ready, please press the space bar.

You will then see a short countdown from 3 to 1. After the countdown, a square symbol will appear, indicating that the clock has started. You now have exactly 30 seconds available for clicking. Remember to use the LEFT mouse button for clicking.

When the time is up, the current round ends. A new screen will appear, showing 'end of round'. After a short moment, you will see the screen with the "+" symbol again, and the next round begins.

There will be 12 rounds in total, and 4 different levels, meaning that you will perform 3 rounds for each level. The sum from all rounds will constitute the amount that will be donated to the charity of your choice.

Important: The total amount of your donation and the charity you choose will be kept confidential. Likewise, you will not be informed of anybody else's total donation.

Important: You cannot earn money for yourself in this part of the study.

We will ask you to put on the headset on your desk and listen to classical music during the clicking phase. This is to minimise the distraction from other participants.

Practice rounds

You will do 4 practice rounds before you begin with the actual task. You can use these rounds to try out which clicking technique works best for you.

All the different levels will appear once, so you can get used to the task.

Please note: No money will be generated for the charity in the practice rounds.

Control Questions

1. Do your earnings for this part depend on the number of clicks?

   □ yes □ no

2. Does your donation to the charity depend on the number of clicks?

   □ yes □ no

3. Suppose in one round you clicked 50 times and each click is worth 3 points. How many points did you donate to the charity in that round?

   …….. points

4. Suppose in one round you clicked 50 times and each click is worth 10 points. How many points did you donate to the charity in that round?

   …….. points

5. Fill in in the gaps.
   Each click can be worth …… , ……. , …….. , or ……. points

6. How many rounds are there in this part?
   …….. rounds

7. How many seconds does one round last?
   …….. seconds

Part II: Click for Charity

Part II is another "Click for Charity" event. It is very similar to Part I. As in Part I, the more clicks you make, the more we will donate to the charity you have already chosen. Again, you cannot earn money for yourself in this part of the study. The differences to Part I are explained below.

Important: When everybody is gathered around to receive their payment, the **two** individuals who produced the **highest donations** in this part will sign their receipts on **yellow paper** and will receive their payment in **yellow envelopes**. Everybody else will sign receipts on white paper and receive their payment in white envelopes.

Furthermore, the names of the two individuals who produced the highest donations will be **highlighted** on the participant list. As mentioned earlier, this list will be **publicly viewable** in the corridor outside of this room. The list will not include any other information on the amounts donated, nor which charities have been chosen.

Only the donations achieved in Part II will count towards determining the individuals who produced the two highest donations.

Instructions for the Second Study

The second study is a "Click for Yourself" event. The task is the same as in the first study. However, the more clicks you make, the more points you will **earn for yourself**. Thus, no charity will benefit from your performance, only you.

The points will be translated into money for **you** according to the same exchange rate as in the first study:

100 points = 2 pence

Hence, your total payment will be calculated as follows:

Your payment = £8 (show-up fee from the first study) + any amount earned in the second study

Your earnings in the second study will remain completely confidential: No information will be shared on who earned much or little in the second study, let alone any specific amounts.

Bibliography

Alesina, A., P. Giuliano, and N. Nunn (2013). On the origins of gender roles: Women and the plough. *The Quarterly Journal of Economics 128*(2), 469–530.

Alpízar, F., F. Carlsson, and O. Johansson-Stenman (2008). Anonymity, reciprocity, and conformity: Evidence from voluntary contributions to a national park in costa rica. *Journal of Public Economics 92*(5–6), 1047–1060.

Alpízar, F. and P. Martinsson (2013). Does it matter if you are observed by others? evidence from donations in the field. *The Scandinavian Journal of Economics 115*(1), 74–83.

Anderson, C. and G. J. Kilduff (2009). The pursuit of status in social groups. *Current Directions in Psychological Science 18*(5), 295–298.

Anderson, C. M. and L. Putterman (2006). Do non-strategic sanctions obey the law of demand? the demand for punishment in the voluntary contribution mechanism. *Games and Economic Behavior 54*(1), 1–24.

Andreoni, J. (1988). Privately provided public goods in a large economy: The limits of altruism. *Journal of Public Economics 35*(1), 57 –73.

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal 100*(401), 464–477.

Andreoni, J. (2006). Philanthropy. In S.-C. Kolm and J. M. Ythier (Eds.), *Applications*, Volume 2 of *Handbook of the Economics of Giving, Altruism and Reciprocity*, pp. 1201–1269. Elsevier.

Andreoni, J. and R. Petrie (2004). Public goods experiments without confidentiality: a glimpse into fund-raising. *Journal of Public Economics 88*(7–8), 1605 –1623.

Ariely, D., A. Bracha, and S. Meier (2009). Doing good or doing well? image motivation and monetary incentives in behaving prosocially. *American Economic Review 99*(1), 544–555.

Beeman, E. A. (1947). The relation of the interval between castration and first encounter to the aggressive behavior of mice. *The Anatomical record 99*(4), 570–570.

Bénabou, R. and J. Tirole (2006). Incentives and prosocial behavior. *American Economic Review 96*(5), 1652–1678.

Bó, P. D., A. Foster, and L. Putterman (2010). Institutions and behavior: Experimental evidence on the effects of democracy. *American Economic Review 100*(5), 2205–2229.

Boksem, M. A. S., P. H. Mehta, B. V. den Bergh, V. van Son, S. T. Trautmann, K. Roelofs, A. Smidts, and A. G. Sanfey (2013). Testosterone inhibits trust but promotes reciprocity. *Psychological Science 24*(11), 2306–2314.

Bolton, G. E. and A. Ockenfels (2000). Erc: A theory of equity, reciprocity, and competition. *American Economic Review 90*(1), 166–193.

Bos, P. A., D. Terburg, and J. van Honk (2010). Testosterone decreases trust in socially naïve humans. *Proceedings of the National Academy of Sciences 107*(22), 9991–9995.

Brañas-Garza, P., A. M. Espín, F. Exadaktylos, and B. Herrmann (2014, 08). Fair and unfair punishers coexist in the ultimatum game. *Scientific Reports 4*, 6025.

Brañas-Garza, P., J. Kováří, and L. Neyse (2013). Second-to-fourth digit ratio has a non-monotonic impact on altruism. *PLOS ONE 8*(4), 1–10.

Buck, J. L. (1935). *Land utilization in China*, Volume 2. Chicago, IL: University of Chicago Press.

Carlsson, F., H. He, and P. Martinsson (2013). Easy come, easy go. *Experimental Economics 16*(2), 190–207.

Carpenter, J. P. and E. Seki (2006). Competitive work environments and social preferences: Field experimental evidence from a japanese fishing community. *Contributions to Economic Analysis & Policy 5*(2).

Carver, C. S. and T. L. White (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The bis/bas scales. *Journal of Personality and Social Psychology 67*(2), 319–333.

Charles, K. K., E. Hurst, and N. Roussanov (2009). Conspicuous consumption and race. *The Quarterly Journal of Economics 124*(2), 425–467.

Charness, G. and M. Rabin (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics 117*(3), 817–869.

Chen, M. K. (2013). The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets. *American Economic Review 103*(2), 690–731.

Cherry, T. L., P. Frykblom, and J. F. Shogren (2002). Hardnose the dictator. *American Economic Review 92*(4), 1218–1221.

Croson, R. T. (2007). Theories of commitment, altruism and reciprocity: Evidence from linear public goods games. *Economic Inquiry 45*(2), 199–216.

De Cremer, D. and T. R. Tyler (2005). Am i respected or not?: Inclusion and reputation as issues in group membership. *Social Justice Research 18*(2), 121–153.

DellaVigna, S., J. A. List, and U. Malmendier (2012). Testing for altruism and social pressure in charitable giving. *The Quarterly Journal of Economics 127*(1), 1–56.

Denant-Boemont, L., D. Masclet, and C. N. Noussair (2007). Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. *Economic Theory 33*(1), 145–167.

Dufwenberg, M. and A. Muren (2006). Generosity, anonymity, gender. *Journal of Economic Behavior & Organization 61*(1), 42–49.

Dwenger, N., H. Kleven, I. Rasul, and J. Rincke (2016). Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in germany. *American Economic Journal: Economic Policy 8*(3), 203–232.

Eagly, A. H., W. Wood, and L. Fishbaugh (1981). Sex differences in conformity: Surveillance by the group as a determinant of male nonconformity. *Journal of Personality and Social Psychology 40*(2), 384–394.

Edwards, D. A. (1969). Early androgen stimulation and aggressive behavior in male and female mice. *Physiology & Behavior 4*(3), 333–338.

Eisenegger, C., J. Haushofer, and E. Fehr (2011). The role of testosterone in social interaction. *Trends in Cognitive Sciences 15*(6), 263–271.

Eisenegger, C., M. Naef, R. Snozzi, M. Heinrichs, and E. Fehr (2010, 01). Prejudice and truth about the effect of testosterone on human bargaining behaviour. *Nature 463*(7279), 356–359.

Elliot, A. J. (2006). The hierarchical model of approach-avoidance motivation. *Motivation and Emotion 30*(2), 111–116.

Faillo, M., D. Grieco, and L. Zarri (2013). Legitimate punishment, feedback, and the enforcement of cooperation. *Games and Economic Behavior 77*(1), 271 – 283.

Falk, A. and U. Fischbacher (2006). A theory of reciprocity. *Games and Economic Behavior 54*(2), 293–315.

Fan, J. (2007). *Geography*. People's Education Press.

Fehr, E. and A. Falk (2002). Psychological foundations of incentives. *European Economic Review 46*(4-5), 687–724.

Fehr, E. and S. Gächter (2000). Cooperation and punishment in public goods experiments. *American Economic Review 90*(4), 980–994.

Fehr, E. and K. M. Schmidt (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics 114*(3), 817–868.

Fincher, C. L., R. Thornhill, D. R. Murray, and M. Schaller (2008). Pathogen prevalence predicts human cross-cultural variability in individualism/collectivism. *Proceedings of the Royal Society of London B: Biological Sciences 275*(1640), 1279–1285.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics 10*(2), 171–178.

Fischbacher, U., S. Gächter, and E. Fehr (2001). Are people conditionally cooperative? evidence from a public goods experiment. *Economics Letters 71*(3), 397–404.

Gächter, S. and B. Herrmann (2009). Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment. *Philosophical Transactions of the Royal Society of London B: Biological Sciences 364*(1518), 791–806.

Gächter, S., B. Herrmann, and C. Thöni (2010). Culture and cooperation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences 365*(1553), 2651–2661.

Gächter, S., E. Renner, and M. Sefton (2008). The long-run benefits of punishment. *Science 322*(5907), 1510–1510.

Galor, O. and Ö. Özak (2016, October). The agricultural origins of time preference. *American Economic Review 106*(10), 3064–3103.

Glazer, A. and K. A. Konrad (1996). A signaling explanation for charity. *American Economic Review 86*(4), 1019–1028.

Gneezy, U., A. Leibbrandt, and J. A. List (2016). Ode to the sea: Workplace organizations and norms of cooperation. *The Economic Journal 126*(595), 1856–1883.

Gneezy, U. and A. Rustichini (2000). Pay enough or don't pay at all. *The Quarterly Journal of Economics 115*(3), 791–810.

Gray, J. A. (1990). Brain systems that mediate both emotion and cognition. *Cognition and Emotion 4*(3), 269–288.

Guernier, V., M. E. Hochberg, and J.-F. Guégan (2004). Ecology drives the worldwide distribution of human diseases. *PLoS Biol 2*(6), e141.

Guiso, L., P. Sapienza, and L. Zingales (2006). Does culture affect economic outcomes? *The journal of economic perspectives 20*(2), 23–48.

Harbaugh, W. T. (1998). What do donations buy? a model of philanthropy based on prestige and warm glow. *Journal of Public Economics 67*(2), 269–284.

Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, and R. McElreath (2001). In search of homo economicus: Behavioral experiments in 15 small-scale societies. *American Economic Review 91*(2), 73–78.

Henrich, J., J. Ensminger, R. McElreath, A. Barr, C. Barrett, A. Bolyanatz, J. C. Cardenas, M. Gurven, E. Gwako, N. Henrich, et al. (2010). Markets, religion, community size, and the evolution of fairness and punishment. *science 327*(5972), 1480–1484.

Henrich, J., R. McElreath, A. Barr, J. Ensminger, C. Barrett, A. Bolyanatz, J. C. Cardenas, M. Gurven, E. Gwako, N. Henrich, C. Lesorogol, F. Marlowe, D. Tracer, and J. Ziker (2006). Costly punishment across human societies. *Science 312*(5781), 1767–1770.

Hermans, E. J., P. A. Bos, L. Ossewaarde, N. F. Ramsey, G. Fernández, and J. van Honk (2010). Effects of exogenous testosterone on the ventral striatal bold response during reward anticipation in healthy women. *NeuroImage 52*(1), 277–283.

Herrmann, B., C. Thöni, and S. Gächter (2008). Antisocial punishment across societies. *Science 319*(5868), 1362–1367.

Hofstede, G. (1980). *Culture's consequences*, Volume 5. Beverly Hills, CA: Sage.

Holt, C. A. and S. K. Laury (2002, 12). Risk aversion and incentive effects. *American Economic Review 92*(5), 1644–1655.

Hsiao-Tung, F. and C. Chih-i (1945). *Earthbound China: A study of rural economy in Yunnan*. Chicago, IL: University of Chicago Press.

Jackson, M. O. and Y. Xing (2014). Culture-dependent strategies in coordination games. *Proceedings of the National Academy of Sciences 111*(Supplement 3), 10889–10896.

Jacquet, J., C. Hauert, A. Traulsen, and M. Milinski (2011). Shame and honour drive cooperation. *Biology Letters 7*(6), 899–901.

Jeong, E. S., Y. Shi, A. Baazova, C. Chiu, A. Nahai, W. G. Moons, and S. E. Taylor (2011). The relation of approach/avoidance motivation and message framing to the effectiveness of charitable appeals. *Social Influence 6*(1), 15–21.

Jinguang, Z. (2004). *Qin Zhi Yan Jiu*. Shang Hai Guji.

Jones, D. and S. Linardi (2012). Gender differences in reputation concerns: Prosocial behavior in the field & lab. Working paper.

Jones, D. and S. Linardi (2014). Wallflowers: Experimental evidence of an aversion to standing out. *Management Science 60*(7), 1757–1771.

Josephs, R. A., M. L. Newman, R. P. Brown, and J. M. Beer (2003). Status, testosterone, and human intellectual performance stereotype threat as status concern. *Psychological Science 14*(2), 158–163.

Josephs, R. A., J. G. Sellers, M. L. Newman, and P. H. Mehta (2006). The mismatch effect: when testosterone and status are at odds. *Journal of Personality and Social Psychology 90*(6), 999–1013.

Karlan, D. and M. A. McConnell (2014). Hey look at me: The effect of giving circles on giving. *Journal of Economic Behavior & Organization 106*, 402–412.

Kataria, M. and T. Regner (2015). Honestly, why are you donating money to charity? an experimental study about self-awareness in status-seeking behavior. *Theory and Decision 79*(3), 493–515.

Knack, S. and P. Keefer (1997). Does social capital have an economic payoff? a cross-country investigation. *The Quarterly Journal of Economics 112*(4), 1251–1288.

Kroemer, N. B., A. Guevara, I. C. Teodorescu, F. Wuttig, A. Kobiella, and M. N. Smolka (2014). Balancing reward and work: Anticipatory brain activation in nacc and vta predict effort differentially. *NeuroImage 102, Part 2*, 510–519.

Leary, M. R. (1983). A brief version of the fear of negative evaluation scale. *Personality and Social Psychology Bulletin 9*(3), 371–375.

Leibbrandt, A., U. Gneezy, and J. A. List (2013). Rise and fall of competitiveness in individualistic and collectivistic societies. *Proceedings of the National Academy of Sciences 110*(23), 9305–9308.

List, J. A., R. P. Berrens, A. K. Bohara, and J. Kerkvliet (2004). Examining the role of social isolation on stated preferences. *American Economic Review 94*(3), 741–752.

Lutchmaya, S., S. Baron-Cohen, P. Raggatt, R. Knickmeyer, and J. T. Manning (2004). 2nd to 4th digit ratios, fetal testosterone and estradiol. *Early Human Development 77*(1–2), 23–28.

Manning, J. T. (2002). *Digit Ratio: A Pointer to Fertility, Behavior, and Health*. Rutgers University Press.

Masclet, D., C. Noussair, S. Tucker, and M.-C. Villeval (2003, 03). Monetary and nonmonetary punishment in the voluntary contributions mechanism. *American Economic Review 93*(1), 366–380.

Mazur, A. and A. Booth (1998). Testosterone and dominance in men. *Behavioral and Brain Sciences 21*(03), 353–363.

Mellström, C. and M. Johannesson (2008). Crowding out in blood donation: was titmuss right? *Journal of the European Economic Association 6*(4), 845–863.

Moav, O. and Z. Neeman (2010). Status and poverty. *Journal of the European Economic Association 8*(2-3), 413–420.

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics 92*(1-2), 91–112.

Nunn, N. and L. Wantchekon (2011). The slave trade and the origins of mistrust in africa. *American Economic Review 101*(7), 3221–3252.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology 46*(3), 598–609.

Qian, N. (2008). Missing women and the price of tea in china: The effect of sex-specific earnings on sex imbalance*. *The Quarterly journal of economics 123*(3), 1251–1285.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review 83*(5), 1281–1302.

Rand, D. G. and M. A. Nowak (2013). Human cooperation. *Trends in Cognitive Sciences 17*(8), 413–425.

Ray, D. (1998). *Development Economics*. Princeton University Press.

Rege, M. and K. Telle (2004). The impact of social approval and framing on cooperation in public good situations. *Journal of Public Economics 88*(7–8), 1625–1644.

Reinstein, D. and G. Riener (2012). Reputation and influence in charitable giving: an experiment. *Theory and Decision 72*(2), 221–243.

Roberts, R. D. (1984). A positive model of private charity and public transfers. *Journal of Political Economy 92*(1), 136–148.

Samek, A. S. and R. M. Sheremeta (2014). Recognizing contributors: an experiment on public goods. *Experimental Economics 17*(4), 673–690.

Schwartz, R. A. (1970). Personal philanthropic contributions. *Journal of Political Economy 78*(6), 1264–1291.

Soetevent, A. R. (2005). Anonymity in giving in a natural context—a field experiment in 30 churches. *Journal of Public Economics 89*(11–12), 2301–2323.

Talhelm, T., X. Zhang, S. Oishi, C. Shimin, D. Duan, X. Lan, and S. Kitayama (2014). Large-scale psychological differences within china explained by rice versus wheat agriculture. *Science 344*(6184), 603–608.

Tangney, J. P., R. S. Miller, L. Flicker, and D. H. Barlow (1996). Are shame, guilt, and embarrassment distinct emotions? *Journal of Personality and Social Psychology 70*(6), 1256–1269.

Terburg, D., B. Morgan, and J. Van Honk (2009). The testosterone–cortisol ratio: A hormonal marker for proneness to social aggression. *International journal of law and psychiatry 32*(4), 216–223.

Titmuss, R. M. (1970). *The gift relationship. From human blood to social policy.* London: George Alien & Unwin Ltd.

Tuiten, A., J. Van Honk, H. Koppeschaar, C. Bernaards, J. Thijssen, and R. Verbaten (2000). Time course of effects of testosterone administration on sexual arousal in women. *Archives of General Psychiatry 57*(2), 149–153.

van Honk, J., E. R. Montoya, P. A. Bos, M. van Vugt, and D. Terburg (2012). New evidence on testosterone and cooperation. *Nature 485*(7399), E4–E5.

van Honk, J., J. S. Peper, and D. J. Schutter (2005). Testosterone reduces unconscious fear but not consciously experienced anxiety: Implications for the disorders of fear and anxiety. *Biological Psychiatry 58*(3), 218–225.

van Honk, J., D. J. Schutter, P. A. Bos, A.-W. Kruijt, E. G. Lentjes, and S. Baron-Cohen (2011). Testosterone administration impairs cognitive empathy in women depending on second-to-fourth digit ratio. *Proceedings of the National Academy of Sciences 108*(8), 3448–3452.

van Honk, J., A. Tuiten, E. Hermans, P. Putnam, H. Koppeschaar, J. Thijssen, R. Verbaten, and L. van Doornen (2001). A single administration of testosterone induces cardiac accelerative responses to angry faces in healthy young women. *Behavioral Neuroscience 115*(1), 238–242.

van Honk, J., A. Tuiten, R. Verbaten, M. van den Hout, H. Koppeschaar, J. Thijssen, and E. de Haan (1999). Correlations among salivary testosterone, mood, and selective attention to threat in humans. *Hormones and Behavior 36*(1), 17–24.

Van Vugt, M. and C. L. Hardy (2010). Cooperation for reputation: Wasteful contributions as costly signals in public goods. *Group Processes & Intergroup Relations 13*(1), 101–111.

Vesterlund, L. (2006). Why do people give? *The nonprofit sector: A research handbook 2*, 168–190.

Warr, P. G. (1982). Pareto optimal redistribution and private charity. *Journal of Public Economics 19*(1), 131–138.

Wen, B., H. Li, D. Lu, X. Song, F. Zhang, Y. He, F. Li, Y. Gao, X. Mao, L. Zhang, et al. (2004). Genetic evidence supports demic diffusion of han culture. *Nature 431*(7006), 302–305.

Young, D. J. (1982). Voluntary purchase of public goods. *Public Choice 38*(1), 73–85.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics 13*(1), 75–98.