Experimental investigation of collective decision making

Theodoros Alysandratos

Royal Holloway, University of London
Submitted in fullfilement of the requirements for the PhD degree

Acknowledgements

I like to acknowledge my parents for without them, I would not be at all. My sister for being there. Stephanie for pushing me through the finish line. Sotiris and Philip for not letting me give up and Michael for not giving up on me. Maris for his guidance and always insightful comments. Anna, Elena, Nora, Zhou, Evridiki and Neil for their friendship and support.

Declaration of Authorship for Co-Authored Work

If you are presenting partly co-authored work, please indicate below your individual contribution to the thesis.

Name of candidate:Theodoros Alysandratos

Thesis title:Experimental investigation of collective decision making

I confirm that the thesis that I am presenting has been co-authored with:

Xiaoyu Zhou, Michael Naef, Sotirios Georganas, Marcello Sartarelli and Konstantinos Katsanevas.

Within this partly co-authored work, I declare that the following contributions are entirely my own work:

(Here you should indicate, in prcis style, the datasets that you gathered, interpreted and discussed; methods that you developed; complete first drafts that you wrote; content that is entirely your own work; etc. It is often appropriate to organise this statement by chapter)

Chapter 1: A tale of two dimensions: Voting with endogenous information acquisition in the lab

This chapter is authored solely by me. I carried all the work required, including the experimental design, theoretical calculations, simulations, programming the experiment, writing the experimental instructions, running the experimental sessions, analysing the data and writing up the results.

Chapter 2: Behavioural underpinnings of the voting decision in the Greek referendum

In this chapter I had extensive involvement in all the stages. More precisely, I contributed the html code required to run the experiments in all the games except for the car game. In this experiment that included writing the experimental instructions, as well. I participated in designing the experiment in three ways. First, I proposed the use of the particular bargaining game, the dictator games and the particular version of the loss aversion task. Second, I participated in discussions with my coauthors regarding which versions of the risk task, the cognitive reflection test and the ambiguity tesk to administer and how to organise the political questionnaire at the end of the experiment. Third, I ran pilot studies that helped determine the precise order of the tasks. Furthermore, I assisted in running the experimental sessions in September. In addition, I cleaned the data we received from the server. I determined how to analyse the data, conducted the analysis and wrote up the results with the exception of part of the introduction.

Chapter 3: Rice farming and the emergence of cooperative behavior

In this chapter I helped determine the research idea and the experimental design, including what games to administer, in what order and how many times to be repeated. I programmed (in English) the risk aversion, the thinking style test and the punishment of strangers versus friends questions. I programmed the coordination task and contributed code towards the dictator and the ultimatum game. I ran two pilot sessions to help calibrate the payments for the experiment. With respect to the analysis of the data, I suggested using the kolmogorov-smirnov tests for the differences in the distributions and to test for differences in the ways that subjects from each area determined whom to punish and how much. I also helped organise the experimental results. I contributed to the external validity section by discovering the China Family Panel Studies and suggesting the

use of the particular variable. I came up with the idea, downloaded and analysed the data from Wikipedia and the World Values Survey. I wrote up Complementary Evidence section alone. Finally, I contributing in writing parts of the Abstract, the Introduction, the Experimental Design and the Conclusion sections.

Signed:
Date:22/9/2017

Abstract

Many decisions of economic interest are taken by groups. However, comparatively little is known regarding how the decisions of individuals are shaped when they are part of a group. In this thesis I investigate, by means of economic experiments, three such cases. First, I use a lab experiment to test how subjects choose whether to acquire information and how they use that information when they need to cast a vote regarding a group decision. My findings suggest that the subjects see this environment as a game, not as a decision making task, and they largely conform to theoretical predictions at the aggregate level. However, I uncover large individual heterogeneity, which is partly explained by behavioural factors. Next, I turn my attention to investigating what behavioural factors correlate with a real, binary, voting decision. In order to do so I administer a battery of validated experimental tests. I find that one decision is associated with fewer correct answers in the Cognitive Reflection Test, suggesting that these subjects are more impulsive. I find no evidence that risk/loss aversion or distributional preferences correlate with either decision. This finding indicates that within my sample the decision is likely driven by the subjects impulsive tendencies, not underlying preferences. Finally, I use a natural variation in crops cultivation to show that a norm has been created and transmitted intergenerationally in one group but not another of otherwise identical people. I argue that the attitude towards punishment, not social preferences or expectations, is the main channel through which the norm is propagated. I link my experimental results to current behaviour regarding contributions to public goods and volunteering to show that my results have external validity.

Contents

List of Figures

List of Tables

A tale of two dimensions: Voting with endogenouns information acquisition in the lab

Theodoros Alysandratos
Royal Holloway, University of London

Abstract

We report on an voting experiment with two dimensional preferences and endogenous information acquisition. Our subjects were asked to pick one out of two colours. One of the colours was the state of the world and the group was paid only if they guessed it correctly. Departing from the existing literature, the values of the two colours to each subject are not equal to each other. Our model suggests the existence of several uninformed voters and that information acquisition depends positively on the values attributed to each colour. Our results at the aggregate level confirm our theoretical predictions. In addition, we find that loss aversion is a strong predictor of both information acquisition and abstention choices. At the individual level we uncover large and persistent heterogeneity. Our subjects keep on making the same mistakes over time and exhibit no signs of learnings. We also find that they tend to vote more often if they in the past received signals that matched the state of the world. We interpret that as a possible sign of overconfidence. Finally, we show that there is no learning and that political polarisation increases the propensity to vote.

Part I

A tale of two dimensions: Voting with endogenouns information acquisition in the lab

1 Introduction

There is a growing literature in experimental political economy that studies voting decisions in the laboratory. In the present study we wish to examine empirically the relationship between the preferences of the voters and their decisions to buy information and to vote. We base our experiment on the framework developped by Oliveros (2013). This model allows us to examine the relationship of the relative intensity of preferences with the aforementioned decisions. In addition, we examine the effect of polarisation. How will the electorate change its behaviour in the presence of biased voters who are never informed but always vote? Subjects in our experiment become informed only when their relative preferences over the two alternatives are sufficiently close. If they become informed, they vote according to their signal. However, a large portion chooses to stay uninformed. Most of the uninformed subjects vote, although a non-negligible minority abstains. The presence of biased voters causes some subjects to change their behaviour, but it doesn't substantially affect the aggregte amount of information in the electorate. However, we uncover large individual heterogeneity and we show that loss aversion has substantial explanatory power. Overall our results show that if polarisation is manifested as an increase in the preference of one alternative over the other, voters might be more likely to vote. However, the aggregate information in the electorate will probably decrease. If on the other hand polarisation is manifested as an increase in the perception of the proportion of uninformed voters, aggregate information and the percentage of the electorate that votes will not change. Finally, our results show that loss aversion might be an important parameter that explains information acquisition and needs to be explored further.

To be more precise: a committee of three has to guess the correct state of the world. One member of the committee is a preprogrammed player following a publicly known strategy. The other two members have the option to buy an imperfectly informative signal before choosing for which colour to vote or whether to abstain. We call models where each voter earns the same reward from every state of the world, as long as her group guesses correctly, unidimensional. In contrast our model has two dimensions, voters may earn different rewards from each guessing correctly, depending on which one the state of the world. In this framework we show that behaviour may differ substantially. More precisely, we predict near-zero levels of abstention and very high levels of uninformed voting, whereas a unidimensional model would predict high levels of abstention and low levels of uninformed voting.

We experimentally test those predictions in the lab and come up with the following results. The fundamental insights of the model find strong empirical support. The subjects remain uninformed when they should not buy the signal. When they vote, they mostly act as the theory predicts conditional on their informational status. We also uncover two regularities that deviate from the theoretical predictions. First, our experimental subjects invested less that expected in information acquisition. Our regressions indicate that loss aversion is a strong predictor of the failure to buy the signal. Second, we observe non-zero, non-trivial abstention rates among the uninformed subjects. Our analysis shows that abstention is increased when political polarisation is lower. Loss aversion also increases abstention. Having observed a signal that matched the state of the world in a previous round decreases the probability of abstention, perhaps indicating an overconfidence effect. Empirical support for the strict predictions of the model with respect to both information acquisition and voting behaviour is not equally strong.

In our design each committee consists of three voters. The committee needs to guess the correct state of the world using simple majority. Two of the members are human subjects who have private, i.i.d. preferences over the two colours, green and yellow. Before casting their vote, or decid-

ing to abstain, human subjects can buy a costly, imperfectly informative, private signal regarding the state of the world. The third player is a preprogrammed, computerised player with a publicly known, predefined strategy. We administered two treatments varying the strategy of the biased player. In the first treatment (unbiased) the preprogrammed player randomises with equal probability between the two options. In the second treatment (biased) she always votes for yellow. We chose these treatments to study whether the presence of an extremely biased decision maker would alter the behaviour of the subjects. Our simulations predict the presence of an alternative equilibrium with abstention in the second treatment. The added benefit of the biased treatment is that it provides a strategically simpler environment. We find no differences between the two treatments with respect to either signal purchasing or voting behaviour. Nor we uncover any evidence that shows the subjects tried to play the alternative equilibrium with abstention.

The present study builds upon three strands of the literature in experimental political economy. Goeree and Yariv (2011) use an experiment to study the effects of communication and decision making thresholds (majority, two-thirds majority and unanimity) in committee voting. Mattozzi and Nakaguma (2016) examine the effect of public vs secret voting on the the decision of the group and show. Both these papers have voters with different preferences over the alternatives. Our model differs from theirs in the sense that we are not interested in the institutional environment, but we focus on the decision to acquire information, the use of information in the voting stage and the effect of the biased voter in the outcome. Bhattacharya et al. (2014) study the effect of abstention versus compulsory voting.Guarnaschelli et al. (2000) vary the size of the group, the decision rule and the communication of the group to study the decision of a jury. They find limited support for the theoretical model of Feddersen and Pesendorfer (1996). Finally, Bouton et al. (2017) create experimental groups to study the effect of various forms of the unanimity rule with respect to information aggregation and group welfare. Our study deviates substantially from the latter three papers in that they induce homogeneous preferences.

The seminal theoretical work on voting with asymmetric information is by Feddersen and Pesendorfer (1996). In that they develop a unidimensional model of voting with exogenous information acquisition. Battaglini et al. (2010) tested that model in the laboratory. In their experiments they found strong theoretical support for the model of Feddersen and Pesendorfer (1996) when they examine the experimental data at the aggregate level. In their experiment subjects who are uninformed abstain when the preprogrammed player is unbiased and those who are informed always vote sincerely, according to the suggestion of the signal. The behaviour of the preprogrammed player forces subjects to vote while being uninformed to level the playing field. Other papers that have studied committees with exogenous information acquisition are Herrera et al. (2016), Morton and Tyran (2011) and Mengel and Rivas (2017). The aforementioned three papers find that uninformed voters sometimes vote and sometimes abstain, wherever that is allowed. Mengel and Rivas (2017) in particular observe subjects voting against their private signals. This behaviour is practically non-existent in our experiment.

Finally, the last strand of the literature that relates to our experiment is the one that studies endogenous informaton acquisition. Grosser and Seebauer (2016) build a unidimensional model of voting with endogenous information acquisition, while varying the size of their committees. They find uninformed voting which they explain with QRE. In our experiment we find that the problem of cursed voting may be even more pronounced when multi-dimensional preferences are introduced. Bhattacharya et al. (2017) develop a similar model, without allowing for abstention. In line with our results they observe overinvestment in costly information acquisition in large groups and underinvestment in small ones. Other studies that relate to ours are those by Elbittar et al. (ming) and Morton et al. (2013).

Our study adds to the literature in two ways: first, we combine endogenous information acquisition with heterogeneous preferences. Second, we introduce the preprogrammed player who allows us to study the effect of preferences polarisation in the electorate.

The rest of the paper is organised as follows: The theoretical set up is presented in section 2. The experimental design is described in section 3. Our main results are presented in section 10. We present more analysis on learning in section 5. Section 11 summarises the findings and concludes.

2 The model

2.1 The set up

In each voting group there are three voters. The group needs to choose between two colours, green and yellow. The goal for the group is to pick the colour that has been drawn by nature to be the state of the world. Each colour is ex ante equally likely to be chosen by nature. The group decision is determined using a simple plurality rule. In case of a tie, a fair coin is tossed to determine the collective choice.

One of the voters is a partisan who follows a predetermined and publicly known strategy. In the unbiased treatment the partisan randomises with equal probability between voting for yellow and voting for green. In the biased treatment she always votes for yellow. In the experiment the partisan voter is represented by a computerised player. In the experiment the partisan voter is played by the computer. We chose this design for two reasons. First, a preprogrammed player is guaranteed to follow the strategy we expect her to play. Since we do not want to study whether the subjects guess what is the optimal strategy for the partisan voter, having a human play that role would introduce unnecessary noise. The second reason for this choice is practical: a preprogrammed player allows us to collect more observations from human subjects in the condition of interest.

The non-partisan voters have to choose from a set of three actions: $\{G, Y, A\}$, where G is vote for green, Y is vote for yellow and A is abstain. Each non-partisan voter earns a private value if the group decision matches the state of the world. Those private values are defined as: $V = \{V_Y, V_G\} \in \{0, 1, \ldots, 100\}^2$. Each private value is independently and identically drawn

from a uniform distribution. If the group picks yellow (green) and yellow (green) is the state of the world, each non-partisan voter will earn her private V_Y (V_G). If the group's choice does not match the state of the world, the non-partisan voters earn zero. Please note that the earnings of each non-partisan voter depend on the group choice, her own choice is relevant only in so far as it influences the group choice.

The non-partisan voters can buy an informative signal $S \in \{S_Y, S_G\}$ regarding the state of the world. S_Y (S_G) means that the signal advises the voter to pick Yellow (Green). The signal costs ten points and its precision is eighty percent; that is the signal matches the state of the world with an eighty percent probability. The precision of the signal is unaffected by the state of the world. Each signal is private and ex ante uncorrelated with the signals that other non-partisans may receive. The distribution from which the private values are drawn, the strategy of the partisan voter, the precision and the cost of the signal are common knowledge.

The timing of the game is as follows: 1) Nature draws a colour to be the state of the world. 2) Each non-partisan voter becomes informed about her own private values. 3) The non-partisan voters decide whether to buy the signal. 4) The private signals are revealed to those who bought them. 5) All voters, partisans and non-partisans alike, cast their votes. 6) The votes are counted and the colour with the most votes is the group choice.

The pure strategy of each non-partisan consists of two decisions: 1) whether to buy the signal or not $I^i : \{1, \ldots, 100\}^2 \to \{\frac{1}{2}, 0.8\}$ and 2) what action to choose in the voting stage $V^i : \{1, \ldots, 100\}^2 \times \{S_Y, S_G, S_\oslash\} \to \{G, Y, A\}$.

2.2 Equilibria

We focus our attention to pure strategies and symmetric, bayesian equilibria. The intuition for the strategies played is the following: Each non-partisan voter has to decide whether to buy the signal or not and how to act upon receiving it or without it. Given that the strategy of the partisan

voter is public knowledge, a non-partisan only needs to consider what the other non-partisan is going to vote in order to determine her own strategy. In other words she has to take into account what the other non-partisan is going to do for every pair of V_Y and V_G values the other non-partisan has. Clearly this is a very demanding calculation. Fortunately, it can be reduced to a simpler one: consider what a non-partisan with a random pair of values will do. Since we are focusing on symmetric equilibria, this line of thought implies that in order to find an equilibrium, it is enough to ensure that both partisans have the same probability of choosing each strategy.

Each voter has to consider the expected payoff of twelve strategies:

1. Vote yellow and remain uninformed:
2. Vote green and remain uninformed:
3. Abstain and remain uninformed:
4. Become informed and always vote yellow:
5. Become informed and vote yellow if the signal is yellow, else abstain:
6. Become informed and always vote green:
7. Become informed and vote green if the signal is green, else abstain:
8. Become informed and always vote as the signal suggests:
9. Become informed and always vote against the signal's suggestion:
10. Become informed and always abstain:
11. Become informed and vote green if the signal is yellow, else abstain:
12. Become informed and vote yellow if the signal is green, else abstain:

It is easy to see that some of those strategies are dominated. For example, it never makes sense to buy the signal if the subject intends to abstain upon receiving it or to vote against the signal. After eliminating the dominated strategies, we are left with strategies (1)-(3), (5) and (7)-(8), to

consider. Following Oliveros (2013) we find the equilibrium in the space of probabilities. That is for every pair of values and a belief regarding the strategy of the other player, we calculate the optimal strategy for player i. The equilibrium is when the probability that player i chooses strategy j is equal to the probability that player $-i$ chooses strategy j. In other words, at the equilibrium both players should have the same probability of voting yellow and remaining uniformed, voting green and remaining uninformed, etc. In order to compute the equilibrium we use computer simulations. It is important to note that it is possible there are other symmetric equilibria that our simulations have failed to uncover. In fact Oliveros (2013) states that equilibrium multiplicity is to be expected in this game.

Result 1 *At the equilibrium there are three main*[1] *types of behaviour: Vote yellow without being informed, vote green without being informed and become informed and vote as the signal suggests.*

The intuition for this result is easy to grasp. Suppose that one's values for yellow and green are 75 and 10 respectively. If that voter bought the signal, her new values, after deducting the cost of buying the signal, would be 65 and 0. In a situation like this, there is no benefit from voting green. Even if it turns out to be the correct state of the world, the voter will earn nothing. She is better off voting for yellow, even if the probability of that being correct is a lot smaller. If the voter knows that she will be better off not using the information of the signal, she is not going to buy it in the first place and she will vote for the more preferred colour, in this case for yellow. Assume now that one's values for yellow and green are 90 and 70 respectively. If she buys the signal, her values are going to be 80 and 60 respectively. However, in this case the difference between the two values and the level of the accuracy of the signal are such that the voter is better off voting for the colour indicated by the signal.

[1] In both treatments in less than 1% of the potential pairs of values it is an equilibrium to always abstain. In the biased treatment it is an equilibrium in less than 1% of the potential pairs of values to vote green if the signal suggests it and abstain otherwise. Due to their very minor importance we do not discuss further the implications of those propositions.

Result 2 *The presence of the biased partisan will cause those subjects for which $VG > VY$ to become more informed relative to the treatment with the unbiased partisan.*

Result 3 *The presence of the biased partisan will cause those subjects for which $VG < VY$ to become less informed relative to the treatment with the unbiased partisan.*

The reaction to the presence of the biased partisan is not symmetric. Those who are favoured by the partisan are more inclined to either vote without acquiring information. If the signal suggests to them to vote for green, there is always one vote in favour of yellow in the ballot. Hence, the expected earnings from buying the signal are diminished and not buying it becomes more attractive. In the same vein, those who favour green are now in greater need to acquire more information.

Result 4 *About two thirds of the voters in both treatments will be uninformed.*

The last result emphasises that due to the asymmetric effect of the biased partisan. While some voters will be less motivated to buy the signal, others will be more motivated. The aggregate effect will be nearly identical.

The computational predictions for the biased treatment are visualised in figure 2. Similarly, the computational predictions for the unbiased treatment are visualised in the figure 8.

3 Experimental design

The experiment took place at the laboratory of Royal Holloway, University of London in May 2015 and it was run on z-tree (Fischbacher, 2007). The subjects were recruited from the graduate and undergraduate student pool of Royal Holloway using ORSEE (Greiner, 2004). In total 68 subjects took part in five sessions. Each session lasted about 75 minutes and subjects

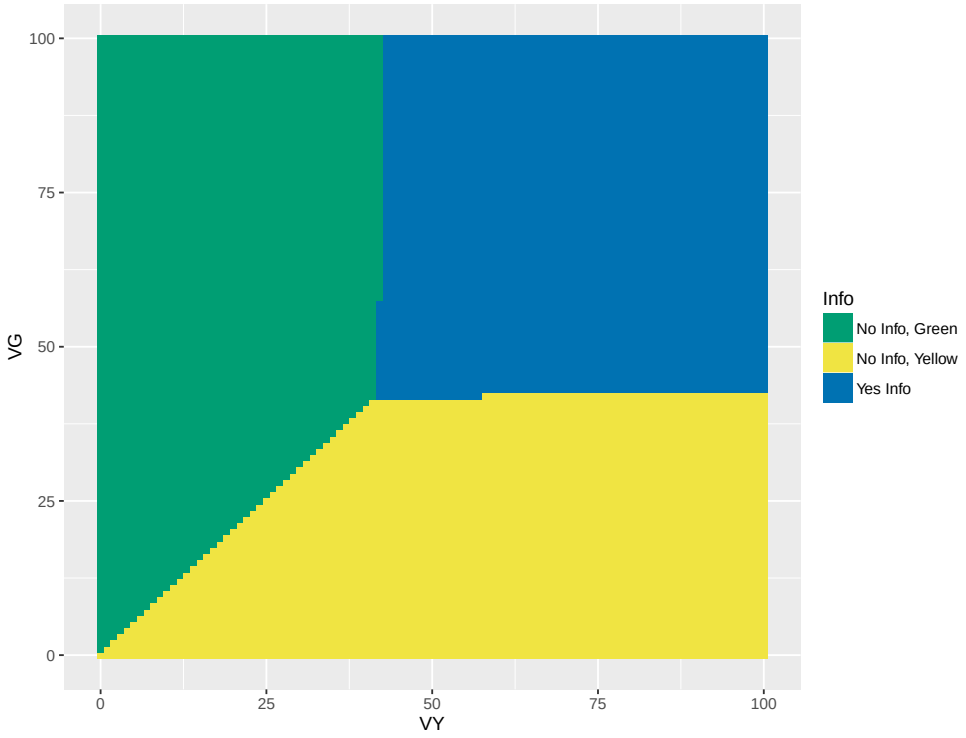


Figure 1: Map of equilibrium behaviour in the unbiased treatment when the other voter is informed.

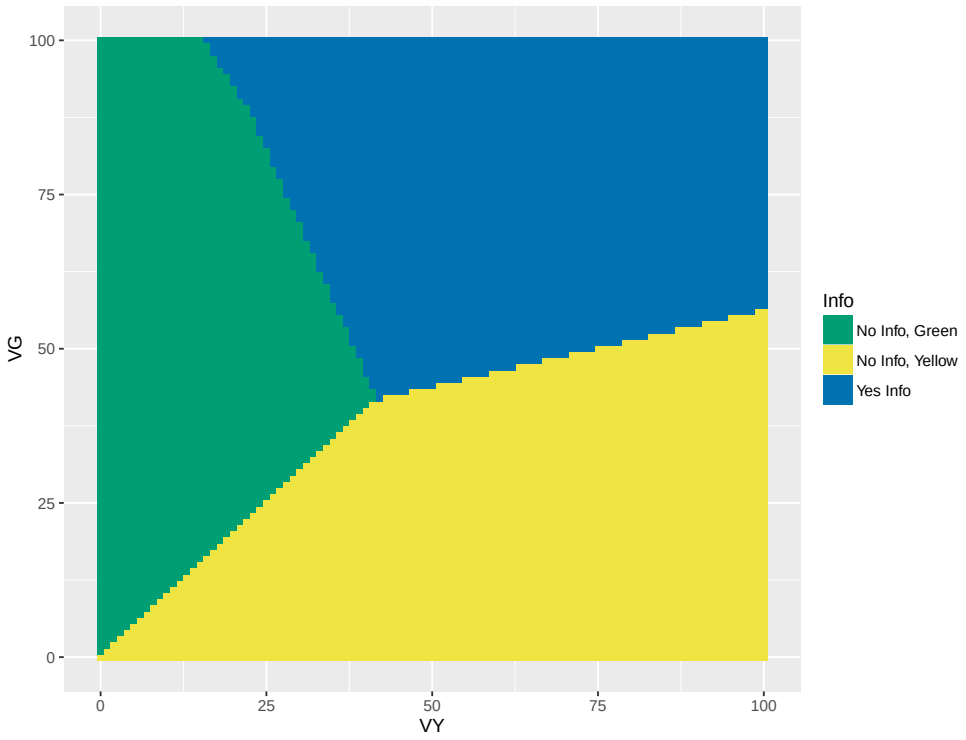


Figure 2: Map of equilibrium behaviour in the biased treatment when the other voter is informed.

Table 1: The order of the treatments and the number of subjects that participated in each of them

Session	Order	Subjects
1	Unbiased-Biased	12
2	Unbiased-Biased	16
3	Biased-Unbiased	20
4	Biased-Unbiased	12
5	Unbiased-Biased	8

got paid for the total number of points accumulated during the experiment. The average compensation, inclusive of a £4 show up fee, was £11.2. We used a within subject design, varying the order of the two treatments to mitigate any order effects. Each treatment lasted for 30 rounds.

The subjects were greeted by the experimenter and detailed instructions were handed out to them. Sufficient time was given to allow the subjects to read the instructions. A questionnaire had to be answered by all the subjects before anyone was allowed to proceed to the next stage. The experimenter also read the instructions aloud to ensure common knowledge of the rules. The wording of the instructions was deliberately kept politically neutral. There was no mention of casting a vote or political candidates so as not to invoke norms with respect to voting. There was no communication among the subjects during the treatments.

During the game the subjects were matched in groups of two in addition to the preprogrammed player, who was computerised, forming a committee of three members. If the group's guess did not match the correct colour, they would earn zero if they had not bought the signal or they would lose ten points if they had. It becomes evident that the subjects could make losses in any given round. This was chosen specifically to instil a sense of real cost regarding information acquisition. Despite that, the cost was kept sufficiently low that no subject could leave the lab without making some positive gains, even in the unlikely event that she made a loss in every single round.

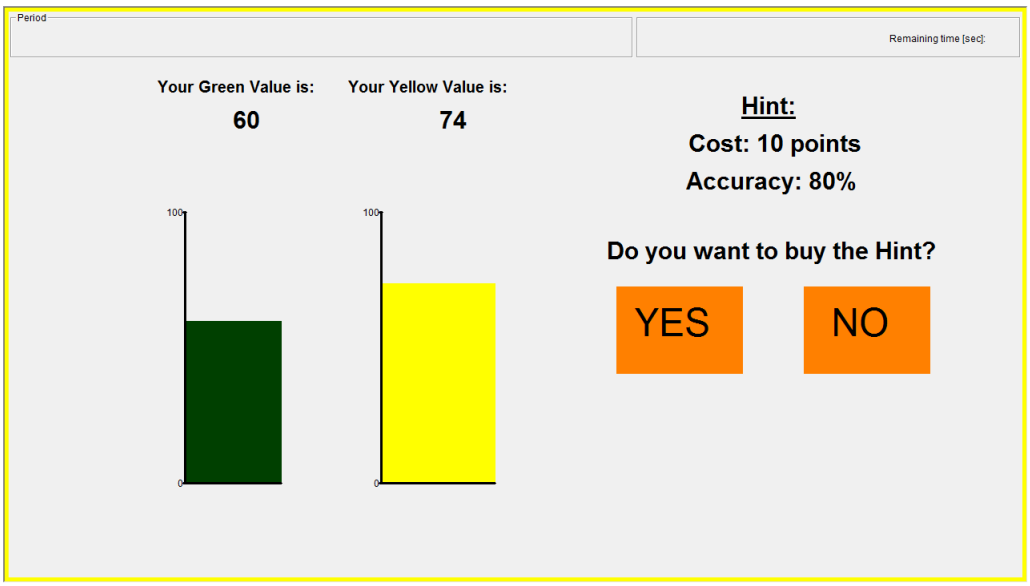


Figure 3: The screen of the subjects in stage 1

In the first stage subjects were shown their private set of values for each outcome. They were also asked whether they wanted to buy the signal. The screen contained a reminder of the cost and the accuracy of the signal. Figure 3 shows the screen the subjects saw at stage 1. Feedback was limited to their own actions in that round, the guess of the group, the true state of the world and the amount of points they earned. We restricted the amount of feedback given to the subjects to prevent, as much as possible, the creation of behavioural norms regarding voting and information acquisition. More precisely, we were concerned that subjects would feel compelled to buy the signal or vote if they knew others did that as well. Similarly we were worried that subjects may punish potential free riders by not buying the signal even if their set of values was sufficiently high to make such a choice non-optimal. Answers to the questionnaire at the end of the experiment point to the direction that the influence of these factors was in fact minimised.

In stage two the subjects were shown the signal, if they had bought it, and were asked to decide whether they would vote in favour of a colour or abstain. Figure 4 shows the screen of the subjects at that stage. Finally, the subjects were given feedback, with the intentional limitations described

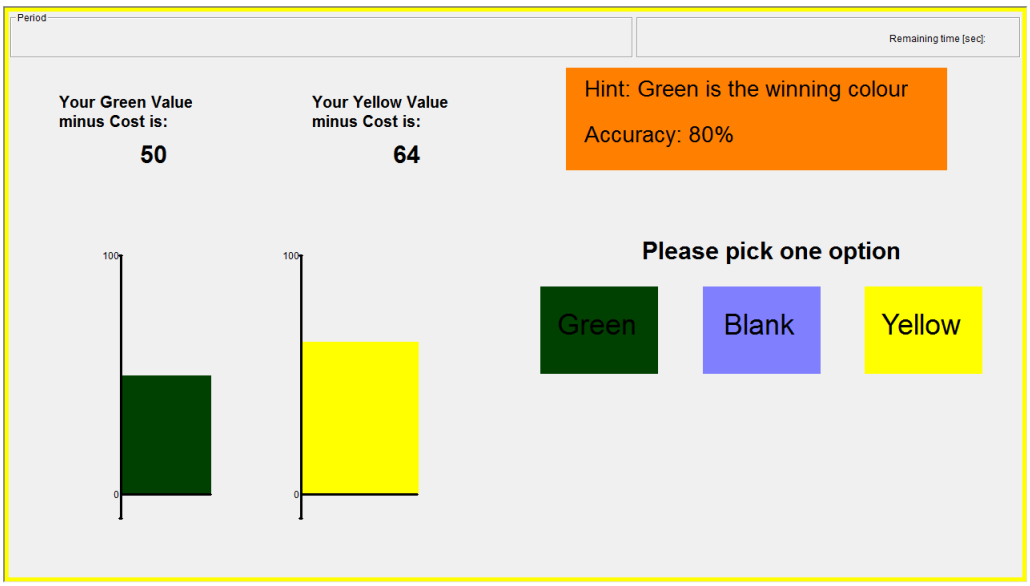


Figure 4: The screen of the subjects in stage 2

above. Figure 5 presents the screen that was used to convey feedback to the subjects.

4 Results

4.1 The decision to acquire information

First, we examine the results at an aggregate level. Our aim is to determine whether the descriptive statistics are in agreement with our theoretical predictions. Figure 6 presents the scatter plot of the decisions made by the subject in the first stage. As it is already evident from the plots, the subjects were largely in line with the theoretical predictions. The mistakes made by those who bought the informative signal while they should not, were close to the border regions. On the other hand, is it more difficult to find a pattern regarding the subjects who should buy the signal but did not based on their values from choosing yellow or green. As can be seen in table 2 83% of the subjects who were expected to not buy the signal, did not buy it. On the contrary, the subjects who were expected to buy the signal, only did so about half the time.

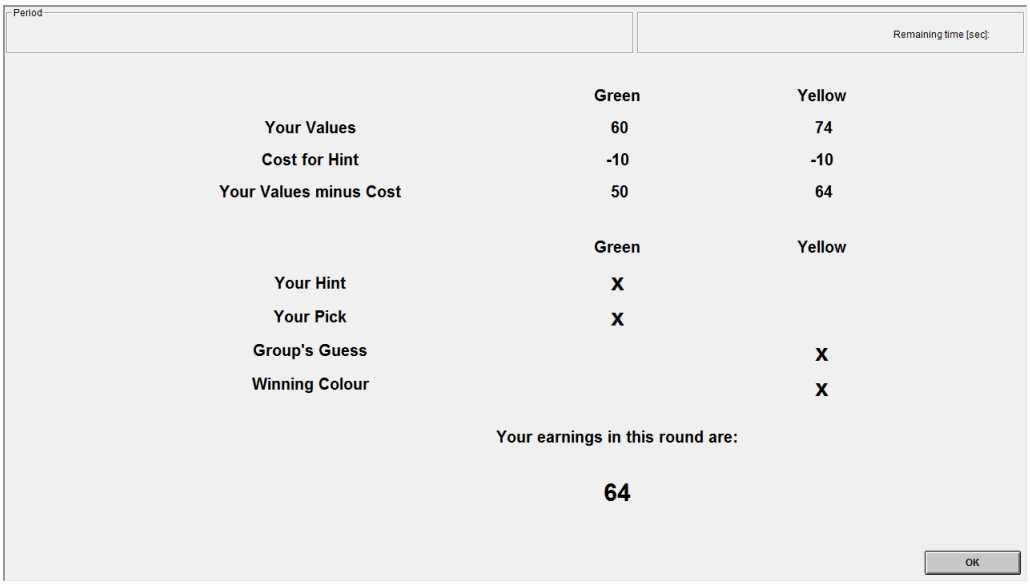


Figure 5: The screen of the subjects in stage 3

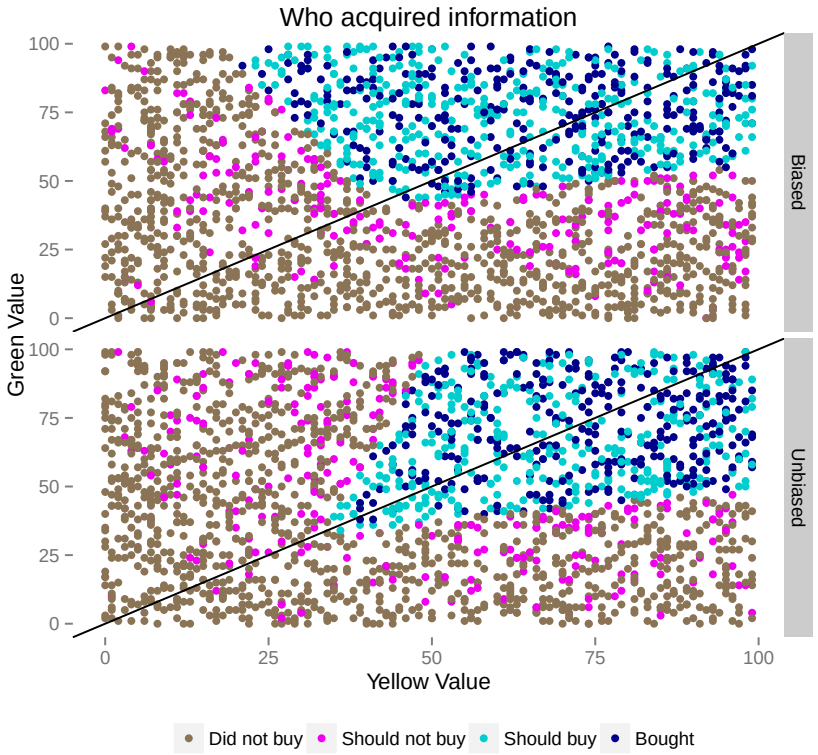


Figure 6: Scatterplot with the decisions of the subjects in the first stage

Table 2: Table of information acquisition decisions made by treatment

	Unbiased Treatment		*Biased Treatment*	
Prediction	Bought	Did not buy	Bought	Did not buy
Buy	48%	52%	45%	55%
Do not buy	17%	83%	17%	83%

Experimental Result 1 *Subjects are more likely to buy the signal as their Yellow and Green values increase.*

Evidence for this result can be seen in table 3. The variables Green Value and Yellow Value correspond to the subjects' Green and Yellow values respectively. Difference is the difference between these two values. Loss Aversion is a binary variable indicating whether the subjects is classified as loss averse or not. Unbiased Treatment is a dummy for the treatment. Total Correct Signals is a variable showing how many times the subjects saw a signal that suggested an action matching the state of the world. The remaining variables are interaction terms between the aforementioned variables, intending to show whether there are differential responses with respect to treatments or colors. The first two columns report results from simple OLS regressions. In column (1) the signs for both colour values are positive and statistically significant. The effect suggests that a one unit increase in Green Value increases the probability of buying the signal by 0.4%. The size effect for the Yellow Value is qualitatively similar. Yet, the two coefficients are statistically different from each other (Wald-test for equality of coefficients, p-value=0.013). In column (2) the sign of Green Value is again positive and statistically significant. More importantly, the Difference variable is statistically significant and positive. The size of the the effect indicates that for a one unit increase in the difference between the two values, the probability of buying the signal increases by 0.4%. This regression indicates that subjects react positively to an increase of the difference when it is due to an increase of the Green Value and negatively when it is due to an increase of the Yellow Value. Column (3) adds period to the list of independent variables in order to see how the decision evolves over time. Our subjects were less likely to buy the signal as time progressed. As we will see later this is unlikely to be due to learning. Col-

umn (4) presents the results of a logit regression that verifies the results of the OLS regressions. Further fixed and random effect regressions, as well as OLS and Logit regressions with clustered standard errors at the session level, that are not reported here, all confirm that the subjects' decision whether to buy or not the signal depends on their values in a way that makes theoretical and intuitive sense.

Table 3: Regressions regarding the decision to acquire information

	Dependent variable:			
	Bought Signal (Yes=1)			
	OLS			*Logit*
	(1)	(2)	(3)	(4)
Green Value	0.004	0.007	0.007	0.030
	p = 0.000***	p = 0.000***	p = 0.000***	p = 0.000***
Yellow Value	0.003			0.023
	p = 0.000***			p = 0.000***
Difference		0.004	0.004	
		p = 0.000***	p = 0.000***	
Loss Aversion	−0.053	−0.054	−0.027	−0.361
	p = 0.00002***	p = 0.00001***	p = 0.019**	p = 0.00002***
Unbiased Treatment	0.005	0.024	0.018	0.072
	p = 0.855	p = 0.040**	p = 0.106	p = 0.790
Total Correct Signals	0.021	0.020	0.029	0.125
	p = 0.000***	p = 0.000***	p = 0.000***	p = 0.000***
Yellow Value*Unbiased	0.001			0.003
	p = 0.127			p = 0.267
Green Value*Unbiased	−0.0002			−0.002
	p = 0.630			p = 0.578
Green Value*Difference		0.00003	0.00003	
		p = 0.00001***	p = 0.00001***	
Yellow Value*Difference		−0.00004	−0.00004	
		p = 0.000***	p = 0.000***	
Period			−0.008	
			p = 0.000***	
Constant	−0.219	−0.172	0.012	−4.711
	p = 0.000***	p = 0.000***	p = 0.546	p = 0.000***
Observations	4,080	4,080	test	4,080
R^2	0.277	0.298	test	

Note:Robust standard errors. P-values are reported in parentheses. *p<0.1; **p<0.05; ***p<0.01

Experimental Result 2 *Subjects are more likely to buy the signal the more times they experience a signal suggesting the correct state of the world.*

Evidence for this result is given in table 3. In all our regressions the sign of the variable Total Correct Signals is positive and statistically significant. The OLS regressions suggest that an one correct signal increases the probability of buying a signal in the future by approximately 2%. This finding indicates that even though our subjects largely conform to the theoretical predictions, they are not immune to behavioural effects. Subjects that have had positive experiences from buying the signal, are more likely to buy it again. This result may help to explain the fall in the consumption patterns of news among countries. In a country where the media have proven to be more accurate lately, voters may be more eager to consume information from them. If, however, the media turn out to be less accurate, even if that inaccuracy is to be occasionally expected, they may be permanently discredited and the voters may turn to other sources.

Experimental Result 3 *Subjects are less likely to buy the signal if they are classified as loss averse.*

At the end of the experiment the following question was posed to the subjects: Assume that someone makes you the following offer: "She will flip a coin. If the coin turns up heads, then you lose 4 pounds; if the coin turns up tails, you win 6 pounds. Would you accept or reject the offer?". The question is taken from Gaechter et al. (2007) and was part of a questionnaire that has been shown to measure loss aversion. The argument in Gaechter et al. (2007) is that subjects would need to have implausibly high risk aversion coefficients to justify declining the wager, therefore it must be loss aversion. We chose only one question in order to maximise the chance of receiving meaningful answers, at the expense of a more refined measure. The subjects who answered they would not accept the bet were classified as loss averse. Our regressions in all columns show a large and statistically significant negative effect of loss aversion on the probability of buying the signal. The size of the effect is more than ten times bigger than that of the value of the colours or of the difference of the values. This result, and especially its magnitude, shows that behavioural factors, such as loss aversion, cannot be ignored by either theorists or policy makers.

4.2 The decision to vote

We now turn our attention to the decision to vote. Table 4 shows how the subjects behaved on aggregate conditional on the information they had. In both treatments informed subjects followed their signal about 95% of the time. We take this percentage as a sign that the subjects understood the game. Given that the signal has an 80% probability of being correct, if one has invested in it, it only makes sense to follow it. The voting decisions of the uninformed subjects are more interesting. About two thirds of them voted for the colour with the highest value. This behaviour is, again, in line with the theory and shows the existence of uninformed voters. Since the decision regarding which colour to vote is clearly driven by either the signal, for the informed subjects, or the colour with the higher value, for the uninformed ones, we focus our investigation to the drivers of abstention. As in the previous subsection, we report results from OLS and Logit regressions with robust standard errors, but fixed and random panel regressions with errors clustered at the session level yield similar results.

Table 4: Table of voting decisions made by treatment

Unbiased Treatment			
Informed		*Uninformed*	
Followed the signal	95%	Colour with highest value	63%
Against the signal	3.4%	Opposite colour	13.5%
Abstained	1.6%	Abstained	23.5%
Biased Treatment			
Informed		*Uninformed*	
Followed the signal	94%	Colour with highest value	68.4%
Against the signal	3.3%	Opposite colour	13.2%
Abstained	2.7%	Abstained	18.4%

Experimental Result 4 *Subjects are less likely to abstain the higher their Green and Yellow values are.*

The evidence for this result is provided in table 5. Abstention is a binary variable that takes value 1 if the subject abstained and 0 otherwise. Bought Signal is also a binary variable, that shows whether the subject bought the signal in the previous stage (1) or not (0). The remaining variables in the

Table 5: Regressions regarding abstention decisions

	Dependent variable:		
	Abstention		
	OLS		*Logit*
	(1)	(2)	(3)
Green Value	−0.001	−0.001	−0.005
	p = 0.000***	p = 0.000***	p = 0.000***
Yellow Value	−0.001		−0.004
	p = 0.000***		p = 0.000***
Difference		−0.001	
		p = 0.000***	
Loss Aversion	0.047	0.047	0.426
	p = 0.000***	p = 0.000***	p = 0.000***
Unbiased Treatment	0.015	0.003	0.099
	p = 0.000***	p = 0.000***	p = 0.0004***
Total Correct Signals	−0.003	−0.003	−0.036
	p = 0.000***	p = 0.000***	p = 0.000***
Bought Signal	−0.123	−0.123	−1.976
	p = 0.000***	p = 0.000***	p = 0.000***
Yellow Value*Unbiased	−0.0002		−0.001
	p = 0.000***		p = 0.00000***
Green Value*Unbiased	−0.0001		−0.0005
	p = 0.000***		p = 0.070*
Difference*Unbiased		−0.0001	
		p = 0.000***	
Constant	0.223	0.229	−1.148
	p = 0.000***	p = 0.000***	p = 0.000***
Observations	4,080	4,080	4,080
R^2	0.058	0.058	

Note:Robust standard errors. P-values are reported in parentheses. *p<0.1; **p<0.05; ***p<0.01

regressions are the same as in the preceding analysis. The signs for the Green and Yellow Value variables are negative and highly statistically significant. In column (2) the sign of the variable Difference is also negative and highly statistically significant.

Experimental Result 5 *The presence of a biased voter makes abstention less likely.*

The sign of the Unbiased Treatment variable is positive and highly statistically significant. However, the interaction terms that involve the Unbiased Treatment are negative and also highly statistically significant. We take this finding to mean that increased polarisation in an election, as is the case in the biased treatment, increases the propensity of the electorate to vote. However, the sensitivity of the voting decision to the value of the colours is diminished in the presence of polarisation.

Experimental Result 6 *Loss averse subjects are more likely to abstain.*

Table 5 shows that the sign of Loss Aversion is positive and highly statistically significant. Furthermore, the size of the effect due to loss aversion is an order of magnitude larger than that due to the magnitude of the values of the colours and at least three times larger than that of the Unbiased treatment. We interpret this to mean that the subjects still perceive the voting decision as bet involving a loss and avoid taking it.

Experimental Result 7 *Subjects are less likely to abstain the more times they experience a signal suggesting the correct state of the world.*

This result is supported by the coefficients of the variable Total Correct Signals. These are all estimated to be negative and highly statistically significant. We interpret this to mean that our subjects develop a sort of overconfidence as they accumulate positive experiences.

5 Learning and decision making

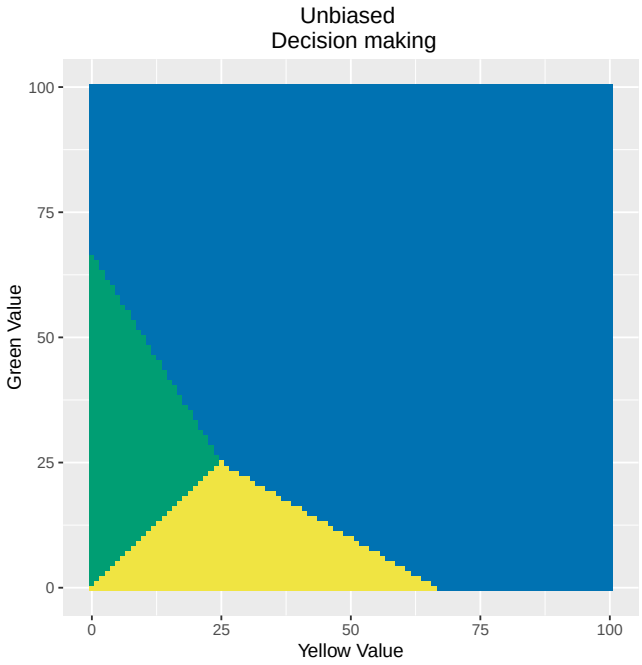
There are a few more points that are worth investigating. Fist, it has been suggested in the literature by Matsusaka (1995) that voters may not perceive elections as a game but, rather, as a decision making task. In our framework it is simple to investigate this supposition. We assume that in such a case the subjects would ignore the presence of the other human player, but not the strategy of the preprogrammed player. This assumption makes sense since the preprogrammed player's strategy is publicly known. Then, we calculate the optimal actions computationally, in the same way that we calculated the equilibrium previously. Figures 7 and 9 show that if this situation was perceived as a decision making task, the number of subjects who would buy the signal would have been a lot higher. Given that the comparative statics only match very remotely and that we already observe lower signal purchasing rates than predicted, we conclude that the subjects perceived the experiment as a game and not as a decision making task.

With regard to learning, first let us look at figure 9. We define two types of errors: acquiring the signal when it is not optimal to do so (positive error) and not acquiring the signal when it is optimal to do so (negative error). In figure 9 we observe that the subjects seem to keep on making the same kind of mistake. They other start by making positive or negative errors and continue on the same path throughout the experiment. There appears to be no learning. The increasing lines indicate that the subjects keep on buying the signal while they should not. The decreasing lines indicate that the subjects keep on not buying the signal while they should. In other words, it shows that while on aggregate the observed behaviour matches the theoretical predictions closely, at the individual level there is large heterogeneity.

Experimental Result 8 *There is no learning in our experiment.*

This statement is supported by the results in table 6. The dependent variable is Error, a binary variable indicating whether an error was committed

Figure 7: Map of equilibrium behaviour in the unbiased treatment in decision making.



(1) or not (0). The coefficient of the control variable Period is not statistically significant. We take this to mean that the propensity to commit an error does not change as the subjects became more acquainted with the experiment. In our design feedback was intentionaly restricted mostly to one's own decisions and the outcomes that affected the whole group. We did that in order to minimise the chance of norms created within the experiment, retaliation and other behavioural effects to affect our findings. It is possible, however, that by not giving our subjects feedback on what the other subjects did, we also restricted their ability to learn from the mistakes of others.

6 Conclusions

We have developped an experiment to test how two-dimensional preferences affect endogenous information acquisition and voting decisions in the laboratory. Our subjects were asked to pick one out of two colours. One

Figure 8: Map of equilibrium behaviour in the biased treatment in decision making.

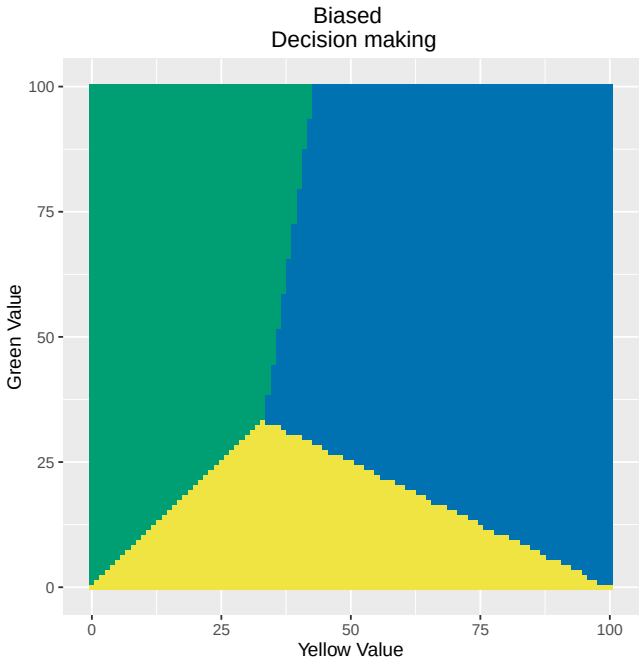


Figure 9: Cumulative errors by session.

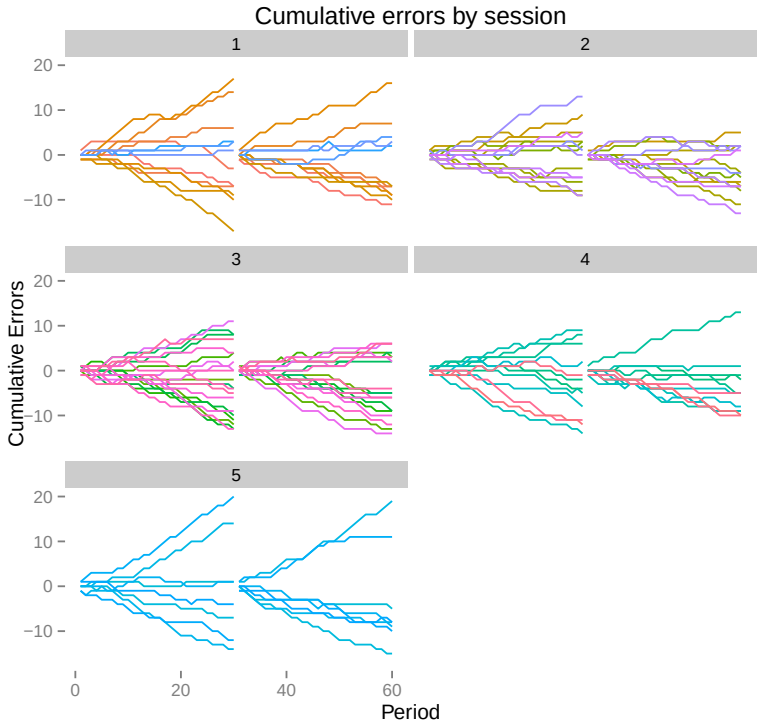


Table 6: Regressions regarding learning.

	Dependent variable:		
	Error		
	OLS		*logistic*
	(1)	(2)	(3)
Green Value	0.005	0.007	0.025
	p = 0.000***	p = 0.000***	p = 0.000***
Yellow Value	0.002		0.013
	p = 0.000***		p = 0.000***
Difference		0.002	
		p = 0.000***	
Period	−0.0003	−0.0003	−0.001
	p = 0.552	p = 0.552	p = 0.556
Bought Signal	0.039	0.039	0.174
	p = 0.097*	p = 0.097*	p = 0.136
Unbiased	0.008	0.008	0.014
	p = 0.765	p = 0.765	p = 0.948
Loss Aversion	−0.031	−0.031	−0.179
	p = 0.022**	p = 0.022**	p = 0.017**
Total Correct Signals	−0.001	−0.001	−0.005
	p = 0.369	p = 0.369	p = 0.425
Yellow Value*Unbiased	0.001		0.005
	p = 0.059*		p = 0.060*
Difference*Unbiased		0.001	
		p = 0.059*	
Green Value*Unbiased	−0.002	−0.001	−0.007
	p = 0.0005***	p = 0.246	p = 0.003***
Constant	−0.019	−0.019	−2.658
	p = 0.456	p = 0.456	p = 0.000***
Observations	4,080	4,080	4,080
R^2	0.110	0.110	

Note:Robust standard errors. P-values are reported in parentheses. *p<0.1; **p<0.05; ***p<0.01

of the colours was the state of the world in each round. Departing from the existing literature, the values of the two colours to each subject are not equal to each other. Our model suggests the existence of several uninformed voters and that information acquisition depends positively on the values attributed to each colour.

Our results at the aggregate level confirm our theoretical predictions. In addition, we find that loss aversion is a strong predictor of both information acquisition and abstention choices. At the individual level we uncover large and persistent heterogeneity. Our subjects keep on making the same mistakes over time and exhibit no signs of learnings. We also find that they tend to vote more often if they in the past received signals that matched the state of the world. We interpret that as a possible sign of overconfidence. We are able to show that the decision making view of the task is not a good match for the observed behaviour. Finally, we show that there is no learning and that political polarisation increases the propensity to vote.

Our findings show that behavioural factors, such as loss aversion and overconfidence brought about by previous positive or negative experiences tend to affect voting and information acquisition decisions. These factors in conjuction with individual heterogeneity need to be taken into account by policy makers when considering how to induce the electorate to become more informed or to vote. Similarly, theoretical models should incorporate these effects in order to more accurately match actual behaviour. Further theoretical and empirical investigation is warranted so as to investigate the interplay of behavioural parameters, such as loss aversion and the game theoretic set up.

Behavioural underpinnings of the voting decision in the Greek referendum

Theodoros Alysandratos[1], Sotirios Georganas[2], Marcello Sartarelli[3], Konstantinos Katsanevas[2]

[1]: Royal Holloway, University of London

[2]: City, University of London

[3]: Universidad de Alicante

Abstract

In the past few decades psychologists and behavioural economists have established that decision makers are boundedly rational. In this paper we investigate whether the proneness of individuals to cognitive biases correlates with actual voting decisions. We administer a battery of established tests in experimental economics measuring risk/loss/ambiguity aversion, depth of reasoning, ability to conduct backwards induction and cognitive reflection. We use the voting decisions in the Greek referendum of 2015 as a convenient test case, as it was a binary choice (Yes or No). We find that "No" voters score significantly lower in the CRT, suggesting they have a lower ability to suppress a spontaneous answer and use System 2 to reason. We also find evidence that "No" voters are strategically myopic, meaning that they fail to grasp the importance of future rounds in sequential move games. On the other hand we find no differences in the distributions of levels in a level-K game. We find no statistical evidence that our subjects differ in their ability to conduct backwards induction, their depth of reasoning, risk/loss aversion and their distributional preferences. These findigns indicate that our subjects have similar preferences in any dimension that we measured. It is therefore likely that their differences in their voting choices was the outcome of higher use of system I cognitive processes, rather than the outcome of different underlying preferences.

Part II

Behavioural underpinnings of the voting decision in the Greek referendum

7 Introduction

Recent referendums in Scotland, Greece and the UK as a whole, whose emergence and consequences have generated a long and intense public debate and are highly relevant to inform policy decisions worldwide, offer illustrative examples. In Scotland voters were asked to decide over independence from Great Britain while, in Greece, over external coordination with other European Union members on the country's fiscal policy and debt. Similarly, in Catalonia, a Spanish region, a coalition of parties running in federal government elections bundled independence from Spain to their elections manifesto, thus implicitly inducing voters to jointly decide over the regional government as well over independence. A referendum on Britain's place in the European Union took place in July 2016. Finally, in the United States, the presidency will be essentially awarded through what amounts to a referendum, between starkly different choices.

This paper investigates major decisions by binary elections in modern democracies: what influences voters? Thinking in these crucial decisions, how prone are they to biases? Do they realize what is good for them on the one hand and society on the other, and can they act upon it? Is it possible to detect common characteristics among the voters who opt for populist candidates, or the policies those candidates support? We answer these questions by using a carefully designed experimental survey that we administered to a sample of in Greece.

Our study is exploratory in nature and does not aim to establish causality. Instead, we wish to uncover potential correlations between behavioural fac-

tors and the decision in the Greek referendum. To that end we recruited a sample of 112 subjects at the University of Athens in July and September 2016. Our subjects are young (average age is approximately 22) and most of them hold an undergraduate degree or study towards one. Their answers in questions that aim to gauge their socio-economic background also indicate no differenfes in that respect. During the experiment we administered a battery of tests that measured risk aversion, loss aversion, depth of reasoning, ability to conduct backwards induction, distributional preferences and the tendency to pick intuitive answers (the cognitive reflection test).

Our results show a strong correlation between the performance in CRT and the voting choice in the referendum. More precisely, we find that voting No is associated with about 0.35 fewer correct answers on average. That figure corresponds to a 30% decrease in the correct answers of the average subject. We find no other statistical evidence for the differences in risk or loss aversion, distributional preferences, depth of reasoning or performance in a bargaining game. These results indicate that, at least among our sample, different choices in the referendum are not driven by underlying differences in preferences. On the contrary, such differences may be attributed to different cognitive processes employed by our subjects to make their mind regarding their choice in the referendum. The finding that CRT is the only variable with explanatory power is significant for those who study the rise of populism and the failure of experts, despite overwhelming majorities among them, to convince the electorate. Voters who are prone to System I thinking may be more easily swayed by populist options that have the tendency to be simple and intuitive than by thoughtful and nuanced arguments that require engagement of System II. Consequently, the outcome of the vote depends on the proportion of intuitive thinkers in the electorate.

Considering the vast number of games administered we skip the traditional literature review here. Instead we are going to discuss relevant papers in the literature when we discuss our experimental design and present results of the respective games. To the best of our knowledge this is the first paper that combines properly incentivised experimental games in order to gauge underlying preferences and cognitive biases and correlates the findings with

choices in the political domain. The remainder of the paper is structured as follows: Section 8 briefly gives the background of the Greek referendum. The experimental design is presented in 9. The results of our experiment are presented in 10. We conclude in Section 11.

8 The Greek referendum

This section is meant to give a brief description of the climate that lead to the Greek referendum and its aftermath. We do not intend this to be an exhaustive account of the Greek financial crisis or a thorough evaluation of the political and economic consequences of the bailout agreement.

Since 2010 Greece has been receiving financial aid from the IMF and European governments in order to finance a vast public deficit that peaked at 15.1% in 2009[2] . In exchange for this financing the country agreed to implement structural reforms. The belief was that those reforms would improve the long term outlook of Greece. Support for this agreement in Greece has been fickle. Political parties have shifted their stance more than once and the public has switched allegiances as well, leading to the rise to prominence of former political pariahs or start up parties, and the fall of former behemoths. The structure of the program has also been criticised as being sub-optimal (Ardagna and Caselli, 2014). The fiscal multipliers used in order to calibrate the program have been questioned (Olivier and Daniel, 2013), too. In January 2015, approximately five years since the country first entered into an agreement for financial support, SYRIZA, one of the upstart parties that previously had single digit support figures, won the elections and formed a government with ANEL, a splinter party formed from Conservative party (ND) dissidents. Both parties had been protesting the financing agreement from the beginning and had formed what was colloquially known as the anti-memorandum front. Their gorvernment promised to renegotiate the deal and redistribute the burdens of the structural reforms in ways that would alleviate those in the most precarious situations.

[2]According to the latest Eurostat data at the time of writing.

In order to do so the government engaged in a highly publicised attempt to improve the financing terms. The culmination of this process was the call of the Greek Prime Minister for a referendum that would ask the electorate to approve or reject the as of then latest offer by the European Union. The Prime Minister announced his decision to hold the referendum in the small hours of 27 June 2015[3]. No prior notice had been given to the members of the Eurozone or the media. Rumours of a referendum had been circulating for months but they had been shut down by ministers and had not gained enough credibility. The governmental proposal was approved by a parliamentary vote the following day. The vote was held on 5 July 2015, only eight days later.

The outcome of the vote was a rejection of the proposal of the European Union by 61% of the electorate. After the announcement of the result the Greek minister of finance and chief negotiator with the European Union and the IMF resigned. On 8 July 2015 the government officially requested a financing agreement from its creditors, promising to implement structural reforms in exchange. On 13 July 2015 Greece and its creditors agreed to a new bailout package that, admittedly, contained larger pension cuts and higher tax hikes than the package that had been rejected by the electorate only a few days previously. After SYRIZA suffered a split of several high-profile politicians who opposed the new agreement and its implementation, the government quit and new elections were called. The vote of 20 September 2015 returned again a majority of SYRIZA-ANEL with only minor losses in their electoral percentage and parliament seats.

It is difficult to find objective answers to political questions. The choices are often informed by the preferences of each individual. This means that reasonable people may disagree. However, in the case of the Greek referendum we can at least surmise an ex post optimal benchmark. This benchmark is given by two facts. First, the government signed a new agreement only days after the results were announced. One would find it hard to argue that any fundamentals of the bargaining process had substantially changed

[3]A number of references to news pieces describing the situation at the time can be found at https://en.wikipedia.org/wiki/Greek_bailout_referendum,_2015

in such a short amount of time. In addition, the new agreement was not an improvement over the old one, that was on offer up until the day the referendum was announced[4]. Second, the electorate validated the decision of the government shortly after in the general elections. This suggests that the majority of the public believed the agreement was at least palatable.

9 Experimental design

The experiment took place over 7 sessions at the Experimental Laboratory of the University of Athens in July and September 2016. We used ORSEE Greiner (2004) to recruit 112 subjects [5], most of them around the age of undergraduate students. The average duration of each session was just under one hour and the average payment each subject received was 10 Euros. The average payment was equivalent to about three and a half hours of work at the legally mandated minimum wage for workers aged under 25. We payed about 0.60 Euros for the CRT plus approximately 9.40 Euros from the variable compensation task. We had a minimum wage of 4 Euros, but all the subjects received more than the minimum.
We administrated a battery of test in the following order: ambiguity aversion test Dimmock et al. (2016), the standard cognitive reflection test Frederick (2005), a loss aversion test Gaechter et al. (2007), the undercutting game designed to elicit levels of strategic sophistication Georganas et al. (2015), a three stage alternating offers ultimatum game following Johnson et al. (2002) [6] designed to measure the subjects' strategic myopia, five three-person dictator games taken from Engelmann and Strobel (2004) to get an indication for the kind of distributive preferences of the subjects, a Holt and Laury (2002) lottery that measures risk aversion, the power to take game Bosman and van Winden (2002) and three rounds of the cars

[4]See for example https://www.bloomberg.com/news/articles/2015-07-13/eu-demands-tsipras-capitulation-as-bailout-costs-spiral-ic1mkgo3 or https://www.politico.eu/article/greece-will-capitulate/. The first article shows that Greece agreed to the demands of its lenders. The second one shows that this agreement was foreseeable.

[5]Eight subjects did not declare the option they supported in the referendum and are therefore excluded from our analysis.

[6]Note that the bargaining game and the power to take game were only administered in September due to technical difficulties.

game, a dynamic chicken game played in real time. No feedback was provided except for the cars game to minimise the effect of learning about the other participants through their actions and preserve, as much as possible, the preconceived beliefs of the subjects. The order of the games was chosen in such a way that the tasks were not repetititive and cognitively taxing tasks were positioned in between easier ones. The decisions regarding the order of the tasks were informed by the feedback of subjects from pilot sessions. The subjects were paid for one randomly chosen game. If a lottery was drawn, they got paid for a randomly chosen selection from that lottery. If the three person dictator game was chosen, they were paid for one randomly chosen dictator game. If the cars game was chosen, they were paid for one randomly chosen round of that game. CRT was incentivised with 0.50 Euros per correct answer on top of the compensation for the variable task. Given the number of tasks administered, we will present the precise design of each task in the section discussing the results in that task. For the games that we will not discuss, we refer the reader to the appendix with the experimental instructions and the screenshots from the experiment.
After the subjects had finished playing all the games, they were asked to fill in a questionnaire that asked them basic demographic data, such as age, gender and level of education, and questions regarding their political behaviour. Among others we asked them what they had voted in the Greek referendum of July 2015 and in the Greek parliamentary elections of Septemeber and January 2015. We chose to ask the political questions in the end so as to keep the subjects agnostic of the aims of our study while they were participating in the experimental tests. Table 7 presents descriptive statistics for the subjects.

10 Results

10.1 An overview

Since we used a barrage of tests, we deem it appropriate to first present the correlation matrix among our main variables. This is shown in 10. As it is evident, the choice in the referendum correlates negatively with

Table 7

Statistic	No Voters		Yes Voters		To
	N	Percentage	N	Percentage	N
Subjects	60	57.7	44	42.3	10
Voted in the referendum	46	76.7	37	84.1	8
In favour of Grexit	23	38.3	1	2.27	2
Women	30	0.5	22	0.5	5
Studying towards or having undergraduate degree	51	85	39	88.63	9
Average age	23.5		22.6		23
Knowledge of the EU (avg. correct answers)	1.25		1.09		1.

the performance in the cognitive reflection test (CRT). There is also weak correlation with the level-K exhibited by the subjects, as well as their risk and loss aversion attitudes. Otherwise, our experimental measures from different games are weakly correlated with each other. Table 8 describes the variables used in the analyses.

Table 8: Description of Variables

Variable Name	**Description**
Referendum (1=No)	The choice of the subject in the referendum.
CRT	The total number of correct answers in the CRT test
Reflective	The subject is classified as reflective if she had at least 2 correct answers in the CRT.
Impulsive	The subject is classified as impulsive if in the CRT she chose the modal error at least twice.
Other	The subject is classified as other if she cannot be classified as either impulsive or reflective.
Level K	The subject's level as inferred from her choice in the Undercutting game.
Bargaining offer	The subject's offer in the first round of the bargaining game.
Bargaining Look Up (1=Yes)	Binary variable indicating whether the subject used the option to look at the size of the pie in the following rounds in the first round of the bargaining game.
Risk	The number of risk averse choices made by the subject.
Loss	The number of loss averse choices made by the subject.

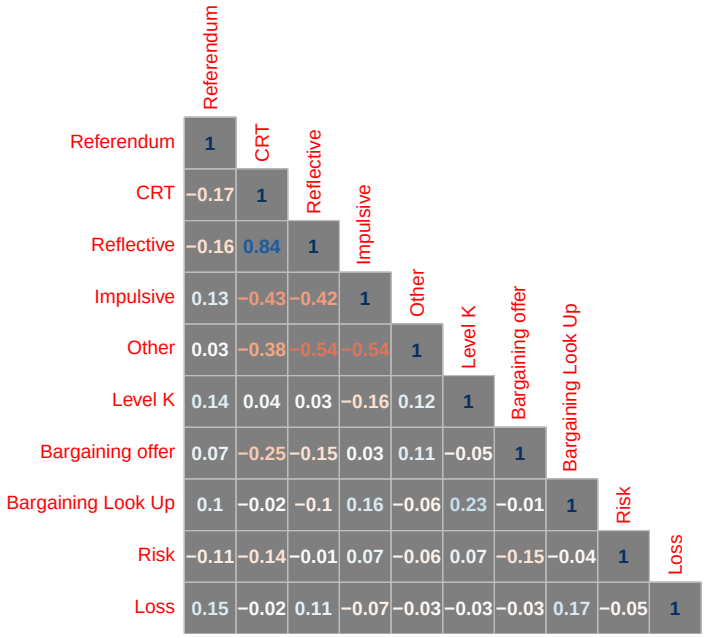


Figure 10: Correlation between the main variables.

10.2 Cognitive reflection test

We administrated the standard CRT Frederick (2005) consisting of three questions:

1. A bat and a ball cost €1.10 in total. The bat costs €1.00 more than the ball. How much does the ball cost?

2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?

3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

Many researchers have noted the existence of two cognitive processes. One process is quick and relies on intuition. The other is slow and involves more deliberate thinking. Those processes are popularly know as "System 1" and "System 2" (Kahneman, 2011). CRT has been shown to reliably capture the tendency to rely on "System 1" when attempting to find the answer

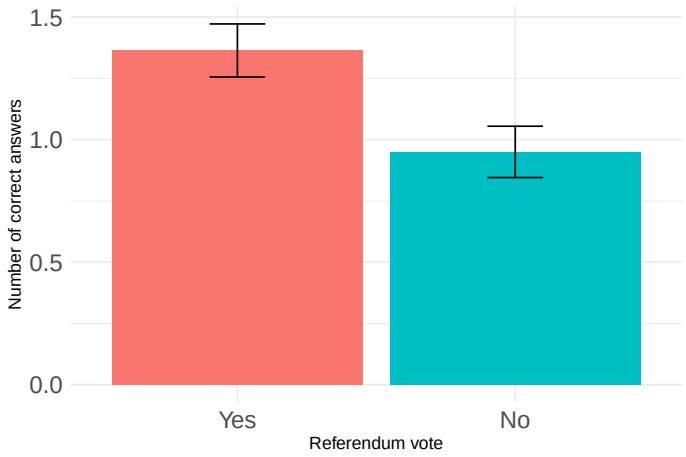


Figure 11: Number of correct choices with error bars in the three questions of the CRT by choice in the Greek referendum of July 2015.

to a puzzle. On many political issues the choice comes down to an option that appears appealing at first glance and one that seems counterintuitive. We do not argue that the appealing option is always correct, however, it may be the case that appealing options have unintended consequences and on further inspection, the counterintuitive answer is preferable.

Result 5 *Subjects who voted Yes in the Greek referendum made more correct choices in the CRT test.*

The subjects who declared that voted Yes gave on average 1.36 correct answers. In contrast, those who declared they voted No gave on average 0.95 correct answers. The difference is highly statistically significant (two-tailed Chi-squared test without Yates' correction, p-value=0.0031) [7] Further evidence for this result are given in table 9, columns 1 and 2. There it is shown that a No vote is associated on average with about 0.36 fewer answers.

In order to further examine the relationship between the vote in the referendum and the performance in the CRT we followed Cueva et al. (2016)

[7]We use a Chi-squared test because both the sum of the rows and the sum of the columns of our 2 × 4 table were not determined before the experiment. Therefore, following the suggestion in Ludbrook (2008) a Chi-squared test is preferred to the Fisher exact test which yields a p-value=0.018. Consequently, the choice of the statistical test does not influence our results.

in assigning our subjects in three categories based on their replies. The subjects that answered at least two questions correctly were classified as Reflective. The subjects who made the modal error in at least two questions were classified as Impulsive. All the rest were classified as Other. The idea is that if a subject gives at least two correct answers, once can with relative safety assume that they are more likely to use System II. On the other hand, if a subject makes the modal error twice, she is more likely to use System I. As can be seen in table 9, voting No in the referendum is correlated positively with Impulsive and negatively with Reflective. However, the p-values of both estimates is just outside the usual levels of statistical significance. Therefore, these regressions at best provide only suggestive evidence regarding the relationship between Reflective (Impulsive) and the vote in the referendum.

Table 9: Results of OLS regressions.

	Dependent variable:				
	CRT		Impulsive	Reflective	Other
	(1)	(2)	(3)	(4)	(5)
Referendum (1=No)	−0.414	−0.359	0.119	−0.125	0.006
	p = 0.009***	p = 0.018**	p = 0.181	p = 0.168	p = 0.948
EU Knowledge		−0.090	0.028	−0.042	0.014
		p = 0.359	p = 0.626	p = 0.479	p = 0.831
Male		0.486	−0.098	0.325	−0.227
		p = 0.002***	p = 0.263	p = 0.0005***	p = 0.022**
Age		−0.021	0.017	−0.003	−0.014
		p = 0.250	p = 0.112	p = 0.766	p = 0.248
Constant	1.364	1.699	−0.195	0.368	0.827
	p = 0.000***	p = 0.0003***	p = 0.457	p = 0.173	p = 0.006***
Observations	104	103	103	103	103
R^2	0.066	0.169	0.064	0.139	0.066

Note: Errors at the subject level. P-values in brackets. *p<0.1; **p<0.05; ***p<0.01

There has been a discussion on whether the proliferation of studies administering CRT is causing the test to lose its efficacy Toplak et al. (2014). A recent metastudy Branas-Garza et al. (2015) finds little support for that conjecture when only laboratory studies are included. Additionaly, our experiment was the first to ever take place at the University of Athens, we know of no other study that has administered CRT in Greece and there seems to be no reason why previous exposure to the CRT is correlated with

the voting decisions of our subjects in the first place. Hence, we believe that CRT is a valid measure of our subjects ability to override their instictive tendencies and provide the correct answers.

10.3 Risk and Loss aversion

One can argue that the decision made in the referendum could be related to attitudes regarding risk and loss aversion. The NO option was inherently more risky, as the range of potential outcomes was, admittedly, wider. On the other hand, YES was a relatively known quantity. In order to test for that assumption we administered the Holt and Laury (Holt and Laury, 2002) measure of risk preferences, which can be seen in table 10.

Result 6 *There is no difference in risk preferences between No and Yes voters.*

The justification for this result can be seen in table 11, in column 1. Risk-Total is the number of risk averse choices made by the subjects. The coefficient is not statistically significant at any conventional level.

Table 10: Risk aversion choices

The following table presents you with ten scenaria. In each of them indicate which one you choose:		
Scenario	A	B
1	€1 with certainty	€10 with probability 50% or €0 with probability 50%
2	€2 with certainty	€10 with probability 50% or €0 with probability 50%
3	€3 with certainty	€10 with probability 50% or €0 with probability 50%
4	€4 with certainty	€10 with probability 50% or €0 with probability 50%
5	€5 with certainty	€10 with probability 50% or €0 with probability 50%
6	€6 with certainty	€10 with probability 50% or €0 with probability 50%
7	€7 with certainty	€10 with probability 50% or €0 with probability 50%
8	€8 with certainty	€10 with probability 50% or €0 with probability 50%
9	€9 with certainty	€10 with probability 50% or €0 with probability 50%
10	€10 with certainty	€10 with probability 50% or €0 with probability 50%

Similarly, one could argue that the reference points of the voters or their sensitivity to loss might differ. Our design controls as much as possible for

the different reference points. Our sample is homogeneous with respect to age, level of education and sex. Additionaly, we asked our subjects to how many countries they have been abroad, as a measure to gauge their financial status. Yes voters had been on average to 5.25 foreign countries, whereas No voters had been on average to 4.27 countries. The difference is not statistically significant (Wilcoxon rank sum test with continuity correction p-value=0.16). We argue therefore that our findings are more indicative of differences in the sensitivity to loss.

Result 7 *There is no difference in loss aversion between No and Yes voters.*

The justification for this result can be seen in table 11, in column 2. LossTotal is the number of risk averse choices made by the subjects. The coefficient is not statistically significant at any conventional level.

Table 11: Risk and loss aversion regressions

	Dependent variable:	
	RiskTotal	LossTotal
	(1)	(2)
Referendum (1=No)	−0.242	0.360
	p = 0.166	p = 0.158
EU Knowledge	−0.096	0.176
	p = 0.400	p = 0.290
Male	−0.123	0.133
	p = 0.476	p = 0.596
Age	0.028	−0.074
	p = 0.197	p = 0.021**
Constant	0.445	5.751
	p = 0.389	p = 0.000***
Observations	103	103
Note: Errors at the subject level. P-values in brackets.	*p<0.1; **p<0.05; ***p<0.01	

Table 12: Loss aversion choices

At first you are given €8. The computer will flip a coin. The outcome is either head or tail. Both outcomes are equally likely. In each of the following scenaria, do you accept or reject the bet?

Scenario	Head	Tail
1	Lose €2	Win €6
2	Lose €3	Win €6
3	Lose €4	Win €6
4	Lose €5	Win €6
5	Lose €6	Win €6
6	Lose €7	Win €6

10.4 Level K and Bargaining

Next, we turn our examination to measures of strategic sophistication. The negotiation process is, at its essence, a bargaining game. For that reason we used the bargaining game in Johnson et al. (2002) to observe the subjects behaviour in a bargaining game which requires the use of backwards induction to calculate the Nash equilibrium. This particular set up also allowed us to investigate what pieces of information the subjects considered relevant in order to formulate their offers.

In sequential games subgame perfections calls for the players to look at the final nodes and start moving backwards in order to calculate the subgame perfect equilibrium. Those of us who have taken a series of courses in game theory may find backwards induction a rather intuitive and simple way of calculating our strategies in a sequential game. However, experimental tests of backward induction typically fail to find strong support for the predictions of the theoretical models (Levitt et al., 2011). Restricting our attention to one-shot games, therefore removing strategic concerns arising from repeatead games, three explanations have been proposed in the literature. The first suggests subjects have other regarding preferences that lead them to take decisions different from the ones predicted by the standard fully rational, self regarding, models. The second is that sophisticated subjects may respond to naive agents. The third one proposes that backward induction is in fact a rather complex and unintuitive process that requires the subjects to consider what will happen at nodes game far from

the one they currently occupy. To make things harder, according to theory one has to consider what will happen at those nodes even though she is not expected to ever reach them. In other words, backward induction asks us to consider future hypotheticals, a task that may not be so easy to the uninitiated.

Take for example a three round, alternating offers ultimatum game in which two subjects, Proposer 1 and Proposer 2, have to divide a shrinking pie. Backward induction predicts that Proposer 1 will look at round three and determine what is the maximum she can make in that round. Then she will consider round two and calculate the maximum Proposer 2 can hope to extract in that round given what Proposer 1 can guarantee herself if she rejects the offer. Finally, in round one Proposer 1 will offer to Proposer 2 the exact amount the latter will get in round two. Assuming both players are national and common knowledge of that rationality, the game ends in round one without ever reaching the following nodes. We administered the experiment described above with one twist: In order for Proposer 1 to gain information for the size of the pie in any round, the current or a future one, she had to click on a tab on her screen. All the tabs were closed by default and at most one tab could be open at any given time. There was no time limit for a subject to make a decision and they could open whichever tabs they wanted, as many times the liked at no cost. When the subjects were ready to make an offer, they would enter a number in a box and press continue. In order to eliminate concerns for social preferences we made Proposer 2 a computerised player, programmed to play the standard, self-interested, subgame perfect Nash equilibrium. and we informed the subjects about its strategy. More precisely, we told them the computerised player wants to make as much money as possible and does not care about the earnings of the human player. In addition, it expects them to try to make as much as possible and it realises human subjects have been informed of its strategy. Our experiment follows very closely one of the treatments in Johnson et al. (2002). The size of the pie in the first round was €20, in the second €10 and in the third €5. Therefore, the subgame perfect Nash equilibrium is for the subject to offer €5 in the first round.

Result 8 *The choice in the referendum does not influence the offer in the bargaining game.*

The justification for this result can be seen in table 13, in column 1. Offer1 is the offer made by the subjects in round 1. Most of the subjects ended the game in the first round, therefore there is no need to analyse offers in further rounds.

Result 9 *Yes voters do not look at the size of the pie in future rounds more often than No voters.*

The justification for this result can be seen in table 13, in column 2. LookUp12 is 1 if the subject in round 1 opened that tab that contained the information regarding the size of the pie in round 2. It is zero otherwise. All of the subjects who looked at the size of the pie in round 3 looked at the size of the pie in round 2 as well. Therefore, there is no need to analyse look ups in rounds other than the second. This finding indicates that No and Yes voters do not differ in their ability to conduct backwards induction.

Various experimental measures of level-K have been used in the literature. We used an one shot normal form game like the ones used in Georganas et al. (2015), called the Undercutting game, which allows for detection of higher levels of depth of reasoning. The payoff matrix of the game we administered can be seen in figure 12. A level-0 player is one who has no clue about this situation and simply picks an action at random. Since a level-0 player would consider all the actions in this game, she would play E with positive probability. E is a dominated action and no other player would choose it. Therefore choosing E is a safe indication that someone is a level-0 player. A player who believes her opponent to be level-0 would be level-1. That player would choose D since that is the best response, assuming the level-0 player randomises among actions using the uniform distribution. A level-2 player would believe her opponent is level-1 and choose C. A level-3 player would choose B and a level-4 player's choice would coincide with that

Table 13: Level K and bargaining regressions

	Dependent variable:		
	offer1	LookUp12	LevelK
	(1)	(2)	(3)
Referendum (1=No)	−0.378	0.102	0.200
	p = 0.911	p = 0.276	p = 0.461
CRTtotal	−6.561	0.053	0.188
	p = 0.011**	p = 0.390	p = 0.289
EU knowledge	1.033	0.048	−0.201
	p = 0.628	p = 0.424	p = 0.242
Male	5.891	−0.040	−0.009
	p = 0.132	p = 0.674	p = 0.973
Age	−0.944	−0.016	−0.008
	p = 0.115	p = 0.155	p = 0.815
LookUp12			0.209
			p = 0.480
RiskTotal	−1.617	0.059	−0.065
	p = 0.371	p = 0.265	p = 0.668
Constant	36.880	0.435	2.109
	p = 0.012**	p = 0.132	p = 0.014**
Observations	60	103	103
Note:			*p<0.1; **p<0.05; ***p<0.01

	a	b	c	d	e
A	0,1 0,1	10 -10	0 0	0 0	-1 1 0
B	-10 10	0 0	10 -10	0 0	0 0
C	0 0	-10 10	0 0	10 -10	0 0
D	0 0	0 0	-10 10	0 0	15 -15
E	0 -11	0 0	0 0	-15 15	-11 - 11

Figure 12: The Undercutting game.

of all higher level players up to Nash and would be A. The subjects were randomly matched with each other and were informed they were playing against another player in the room.

Result 10 *Choice in the referendum does not predict the degree of sophistication.*

The justification for this result can be seen in table 13, in column 3. LevelK is the level of reasoning implied by the choice each subject made in the Undercutting game.

10.5 Distributional preferences

Finally, we administered five three person dictator games in order to get an indication regarding the subjects distributional preferences. Political choices, including those in a referendum, may be driven by preferences regarding the distribution of wealth in the economy. Besides, the government that called the referendum was elected promising to redistribute the burdens of the financing agreement. The games we used are taken from Engelmann and Strobel (2004) and correspond to their Fx, Ex, N, R and Ey treatments. The subjects were divided in groups of three and were asked to make decisions as if they were Person 2. They were told this game was

chosen to determine their payment, then the decision of one of the group's members would be implemented.

In each of these five games a subject can pick one of three distributions. At least one distribution maximises maximin preferences and one maximises efficiency preferences. We diverge from the classification in Engelmann and Strobel (2004) in that we do not explicitly consider preferences that lie in between maximin and efficiency maximisation. Instead, because of the number of times our subjects chose it, we call choice B as indicating a third category, which we call "Other". The games can be seen in table 14

Table 14: Dictator games

	Dictator game								
	Game 1			Game 2			Game 3		
Allocation	A	B	C	A	B	C	A	B	C
Person 1	17	10	9	21	17	13	16	13	10
Person 2	10	10	5	12	12	12	8	8	8
Person 3	9	10	1	3	4	5	5	3	1
Efficiency	A			A			A		
Maximin	A					C	A		
Other		B			B			B	

	Game 4			Game 5		
Allocation	A	B	C	A	B	C
Person 1	11	12	2	21	9	3
Person 2	8	12	3	17	9	4
Person 3	9	12	4	13	9	5
Efficiency	A			A		
Maximin			C			C
Other		B			B	

Result 11 *The distributional preferences do not differ between those who voted Yes and those who voted No.*

In order to test for differences in the distributional preferences we counted the number of times each subject made a choice that corresponded to "Efficiency", "maximin" or "Other. If a subject had made more "Efficiency" choices, was considered an efficiency maximiser. If a subject had made

more "maximin" choices, was considered a maximin player. If the subject had made more "Other" choices, was considered to belong to the other category. There were a 6 subjects that following this classification algorith were not assigned to any of the above categories and were considered uncategorisable. After this categorisation, we performed a Chi-squared test. The test shows no differences at the usual levels (p-value=0.102). However, the p-value is marginal and may be considered suggestive evidence for the presence of such differences.

11 Conclusions

We have administered a battery of tests in order to uncover behavioural reasons that correlated with the vote in the Greek referendum of July 2015. Our experiments took place at the University of Athens and our pool consists of 112 subjects, approximately 22 years of age with an undergraduate degree. Our scope was to investigate the existence of behavioural differences, not to produce a nationally representative sample.

Our tests show that the strongest correlation is between the performance in the Cognitive Reflection Test. In particular, a smaller number of correct choices is correlated with those who voted No. In the literature a low number of correct choices in CRT is associated with impulsivity and increased used of the System I thinking process, which is fast and intuitive. One could then argue that the No option was the more intuitive answer to the question posed to the electorate.

We find only suggestive evidence, that fail to cross the usual thresholds for statistical significance, regarding differences in risk and loss aversion and distributional preferences. All three of them may be considered factors that can influence the choice in any elections. Frequently candidates promote themselves as the stable option, strongly indicating that the status quo is not threatened by them. Or, in contrast, others present themselves as disrupting forces, that come to overturn the establishment. However, in the context of the Greek referendum, we do not find compelling statistical evidence to support the notion that framing the choices as risky or po-

tentially causing a loss swayed the part of the electorate that matched the characteristics of our sample. The same holds for distributional preferences.

Finally, we administered two tests design to gauge the ability of our subjects to conduct backwards induction and to measure their depth of reasoning. Given that election choices might refere to complicated strategic environments, it was worth checking whether the ability of our subjects to reason strategically played a role. We find no evidence that such abilities were correlated with either choice in the referendum.

Rice Farming and the Emergence of Cooperative Behavior

Theodoros Alysandratos[1], Xiaoyu Zhou[1], Micahel Naef[1]

[1]: Royal Holloway, University of London

Abstract

The origins of the observed differences in cooperativeness among societies have puzzled social scientists for years. This paper presents evidence that these origins may be attributed to cultural norms shaped by environmental factors and transmitted intergenerationally, even when the original conditions that lead to their birth are no longer present. In addition we present evidence that the norm transmission operates via the attitude with respect to punishing free riders. In particular, we hypothesise that the cultivation of paddy rice hundreds of years ago influences the decisions made in our incentivised economic games. Unlike cultivating other crops, such as wheat or corn, farming rice historically required extensive cooperation among farmers. Therefore, centuries of rice farming might lead to the creation of a cooperative social norm that affects people living in that society and who transmit it from generation to generation. To test this hypothesis, we travelled to four typical rice and non-rice provinces in China and recruited a total of 524 local undergraduate students as subjects. We find 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, but no difference in how the two groups react to punishment. It follows that the different levels of cooperativeness observed in the punishment treatments are a direct effect of the significant differences in the frequency of punishment. Furthermore, as there is no difference between the two groups in the ultimatum 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. We attribute the differences to a cultural norm resulting from a history of farming that affects the whole population living in that society. We complement our findings using a natural field experiment and two surveys regarding the provision of three different public goods. In all these measures we find evidence of higher provision in rice areas, lending external validity to our experimental results.

Part III

Rice Farming and the Emergence of Cooperative Behavior

12 Introduction

Studies have documented that people's level of cooperativeness differs across countries and societies.[8] However, the reason that these differences exist in the first place 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 a traditional agricultural activity practiced hundreds of years ago gave rise to a cooperative social norm that persisted over generations. We find that this norm has a profound influence on subjects' behaviour in an incentivised and strategic setting. The agricultural activity refers to the cultivation of rice. 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 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 comparison to farming wheat (Buck, 1935). Therefore, families

[8]See Herrmann et al. (2008) and Henrich et al. (2006) for differences in punishing 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.

with only a few labour 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 neighbouring villages to deal with the labour shortage issue. In sum, farming rice historically requires extensive cooperation among farmers and non-farmers, and the hypothesis is that centuries of rice farming leads to a cooperative social norm that is intergenerationally 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 were farmers.[9] Even in 2011, about 34.8% of China's population was employed in the agricultural 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 (1996), because we do not want the farming statistics to be affected by recent technological advances, but rather to reflect, as closely as possible, the historical farming choices. Indeed, the two rice provinces in our sample have been prominent rice production provinces since the Song Dynasty (960 – 1279) (Fan, 2007).

We recruited local, Han Chinese, first year university students based on their Hukou. We obtained data on the subjects' place of origin on their Hukou from the official records kept at their universities. 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

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

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.[10] We also asked the subjects to self-report the place where they grew up. The two measures are highly correlated (Spearman's rho 0.9557, $p < 0.001$), thus we use the place of origin based on their Hukou, as that is the official record.

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 understanding of the differences between rice and non-rice provinces, we also asked subjects to participate in a dictator game (DG), an ultimatum game (UG), and a coordination game.

There are several important features of our experimental design worth mentioning. First and foremost, our subjects were not farmers but university students who had minimal farming experiences.[11] Therefore, the present study goes beyond merely testing the effect of rice farming on farmers' 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 that might have burdened similar studies that relied on students recruited from one Beijing university [12]. The reason is that Beijing universities set small quotas for students from provinces other than Beijing, therefore, non-Beijing students who intend to study there 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

[10]Read the experimental section for a detailed discussion of the recruitment criteria and procedure.

[11]See the experimental design section for a discussion of why this is the case.

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

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 level as non-rice subjects in the last period. On 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.

One might be concerned that these differences are only noticeable in the artificial circumstances of lab experiments. In order to address these concerns we also tested the hypothesis using data from one natural field experiment and two surveys regarding the provision of public goods. The natural field experiment refers to the contributions to Wikipedia. The on-line encyclopedia depends on voluntary contributions for its content and once content has been added on a page it is accessible to anyone, regardless of whether they contributed or not. Therefore, Wikipedia is a prototypical public good. We focus on the Chinese language version in order to minimise noise from

contributors who are not Han Chinese. In order to circumvent the problem of the widespread use of VPN services in mainland China that render the geolocation of contributors via their IP address potentially meaningless, we only used data from the pages of 206 Chinese cities listed on Wikipedia [13]. We argue that thank to the topic of those pages contributors are more likely to be related to the cities and therefore indirectly provide a way to pin down their origins. We find that even when we control for GDP and the popularity of a page, the number of edits per 1000 residents, our measure of contributions, is statistically significantly higher on the pages of cities in rice areas. Furthermore, we use data from the sixth wave of the World Values Surveys. From the answers in the the survey we construct two measures of volunteering, one in the intensive and one in the extensive margin. After controlling for a number of other variables, we find statistically differences in both measures. Finally, we use the Chinese Family Panel Studies survey, a large, representative survey of the Chinese population. The questionnaire contains interviewer's observations on the tidiness of the street of the interviewee. We argue that the tidiness of the street is a local public good. According to our analysis the streets of the interviewees in rice provinces are tidier. Even though our complementary measures taken individually have limitations, taken as a whole they point to actual differences in the willingness to provide public goods in China. In all cases the differences are in line with our experimental results and support the hypothesis we tested.

To the best of our knowledge, this is the first paper that uses experimental methods to show how centuries old agricultural activities have a long lasting and profound effect on individual's level of cooperativeness and punishing behaviour nowadays. We go beyond demonstrating mere differences, investigating the mechanism that gives rise to them. We show that in this case the difference in the probability of punishing free riders is the main channel. Finally, we go beyond the current literature relating our experimental results to various measures of contributions to public goods outside

[13]The list that we used to find those cities can be found on: https://zh.wikipedia.org/wiki/%E4%B8%AD%E8%8F%AF%E4%BA%BA%E6%B0%91%E5%85%B1%E5%92%8C%E5%9C%8B%E5%9F%8E%E5%B8%82%E4%BA%BA%E5%8F%A3%E6%8E%92%E5%90%8D

of the laboratory, demonstrating that these measures support our experimental results. 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, as measured by the Triad task.[14] However, unlike Talhelm et al. (2014), we focus our attention on the economically relevant aspects of behaviour 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. Another closely related paper is that by Soo Hong et al. (2015) who find that the amount of rice paddy fields is associated with higher levels of cooperativeness, as measured by a one shot PGG without punishment. Their experimental design has substantial differences to ours. First, the use of an one shot game without punishment does not allow them to investigate the channel that leads to those differences. In our experiment we conclude that the cultivation of rice has little effect on people's cooperativeness per se, instead the biggest effect comes through teaching the individuals how to use punishment to foster cooperation. We are also able to reject the thinking style and social preferences as the drivers behind our results. Moreover, they conducted their experiments in Beijing. We have already explained why a sample of students recruited from Beijing may not be representative of their original provinces. The fact that the finding of Soo Hong et al. (2015) is in line with ours shows that the effect of the cultural norm we detected may hold even among the part of the distribution to which the highly selected students in Beijing belong. Hence, we view our results as complementary to ours.

This research contributes to the emerging literature that aims to explore the origins of the observed differences in people's preferences across societies. Alesina et al. (2013) investigate the origins of cross-cultural differences in norms regarding gender roles. They find that societies which historically practice plough agriculture have less equal gender norms in the work place.

[14]We use the rice and non-rice terminology instead of rice and wheat. This is because in addition 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.

The rationale behind this observation is that the use of plough requires a considerable level of upper body strength, hence, men have a relative advantage in the workplace in those societies. Galor and Özak (2016) find that participation in agricultural activities affects people's time preferences. In particular, they argue that because the agricultural sector yields higher but delayed returns farmers are more future oriented. Nunn and Wantchekon (2011) find that the level of mistrust within Africa stems from the transatlantic and Indian Ocean slave trades. There is also a small but emerging literature that utilises 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. Carpenter and Seki (2006) also find evidence supporting the idea that work place organization has profound influences on cooperation.

The present paper adds to the literature in several ways. First, we contribute by focusing on explaining the origins of cross cultural differences in cooperativeness, complementing papers that have focused on gender norms, time preferences, and trust. Second, we go one step further identifying the exact channel that gives rise to these differences. In our case it appears to be the use of punishment and not inherent differences in pro-social preferences or strategic behaviour. Third, our subjects were university students with very similar back grounds and have minimal farming experiences. This not only shows that our results are less likely driven by other factors, but also suggests that the social norm created by rice farming is able to influence people even outside the agriculture sector and is stable across generations. This result is in sharp contrast to the findings of Gneezy et al. (2016) and Leibbrandt et al. (2013) who do not detect 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). Finally, we have a better way to control for self-selection or migration into

or out of rice areas that may bias our estimations. We do so by merging our experimental observations with administrative data obtained from the universities.

The remainder of the paper is structured as follows: Section 13 illustrates the experimental design. Non-parametric and regression results are presented in section 14. A series of Robustness checks are ran in section 15. We present complementary evidence from surveys and a natural field experiment in section 16 Section 17 concludes.

13 Experimental Design

The ideal experimental design would compare subjects of identical backgrounds that differ only in the type of cultivation in the area where they were exposed. Our recruiting protocol attempted to get as close to this ideal as practically possible. Another feature of our design is the recruitment of subjects that are not farmers themshelves. We, thus, test for the existence of an intergenerationally transmittable cultural norm, not merely for a norm shaped by ecological factors.

13.1 Province selection

We conducted lab experiments in four universities that are located in four provinces across China. The provinces were chosen from the list of provinces in Talhelm et al. (2014) that had more (less) than 70% (10%) of cultivated land devoted to rice and according to lab availability. By restricting the sample to Chinese subjects, a number of potential confounds, such as language, political institutions and other cross country cultural differences are controlled by design.

Since we are investigating the intergenerational transmission of a social norm brought about by rice farming, the categorization of rice and non-rice provinces should be based on the type of crops farmers historically cultivated. For this purpose, we use the earliest available cultivation data

Table 15: Descriptive statistics table

	Non-Rice		**Rice**	
	Hebei	Shandong	Hunan	Zhejiang
		Subjects		
Female	55.7%	74.3%	57.9%	55.6%
Rural	69.9%	60.5%	59.5%	66.9%
Natural Sciences Stream	67.8%	55.9%	50.8%	71.8%
Only Child	31.3%	41.4%	40.5%	62.1%
Relative Income	1.6	1.9	1.8	2.0
Collectivist Thinking	81.1%	82.7%	84.3%	78.2%
		Provinces		
University Rank	100-120	100-140	130-150	250-300
Minimum Wage (in ¥)	12-15	13-16	10.7-13.5	12.5-17
Percentage of Rice	1.9%	2.3%	78.9%	83.1%
Subjects	116	156	128	124

(1996) from the Bureau of Statistics Web site.[15] Following Talhelm et al. (2014), a province is classified as rice if more than half of its cultivated land is devoted to rice paddy fields. The two rice provinces we chose were Hunan and Zhejiang province, which devoted more than 78% of the cultivated lands to paddy fields. Also note that these two provinces have been prominent rice farming provinces since the Song Dynasty (Fan, 2007). The percentage of rice paddy fields was less than 2.5% in the non-rice provinces: Hebei and Shandong. Table 15 presents descriptive statistics for the subjects.

13.2 Subject Recruitment

We recruited local, Han, first year university students. We argue that our subjects have minimal farming experience. About 40% of the subjects in our sample hold an urban Hukou which implies that their families do not have the legal right to possess farming lands.[16] These students live in urban areas and hence have minimal experience in agricultural activities. The rest

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

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

of our sample are students from rural areas, which implies that their families are entitled to farm lands and hence they might have sufficient farming experience. However, we argue in the following that this is not the case. Firstly, the National Higher Education Entrance Examination (Gaokao) is the most important exam not only for students, but also for the schools, since their ranking depends on the student's performance. Therefore, for the three years of study in the senior high school, they have a very busy schedule in order to succeed in the exam. Secondly, 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 need to live on campus during school days and hence it is not possible to work in the field. One might argue that they can help their families during weekends or holidays. However, we doubt this limited time of farming experience would able to become a major influence on their behavior. Third, rice cultivation nowadays requires less cooperation due to advances in technology, therefore, even if subjects spent sufficient time in rice farming, the norm of cooperation stems from the traditional farming techniques, rather than the modern practice. Fourth, in our analyses we controlled for a subject's Hukou, and the results are not affected. This suggests that even if subjects from rural areas have sufficient farming experience, our results are not mediated by them, instead we are capturing an intergenerationally transmitted norm. Finally, we also tested whether *within* rice regions, subject who have rural Hukou behave differently from subjects who have urban Hukou. We find small differences in this regard. More precisely, subjects from rural rice areas employ more prosocial punishment than their counterparts from urban rice areas. We believe this result shows both the channel through which the norm operates and validates our view that the norm is well established in the regions and is minimally affected by current practices. Thus, we believe that our results are less likely to be mediated by the subjects' personal farming experiences and more likely to be influenced by the intergenerational transmission of the social nom 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, the percentage of farming land devoted to rice and non-rice crops is very close to 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.[17]

We chose Han Chinese because Wen et al. (2004) show that Han Chinese have the same cultural origin. We also recruited first year students because they had just graduated from high school. Chinese high schools have a busy and nationally mandated curriculum: students spend more than seven hours per day in the classroom and study similar material. Therefore, subjects had similar experience prior to university. Moreover, first year students are free from the indoctrination effect of their field of study.[18]

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 staff of each university. We provided a script about how to recruit the students. We emphasized that it was an economic study, they would receive mmonetary reimbursement 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.

[17]When analyzing the data, we find that a few subject's Hukou was not from the province of the experiment. In order to utilize as many observations as possible, we drop subjects whose Hukou was from a province farming a crop different from that of the experimental province. For example, if the experiment was conducted in a rice province, we drop subjects whose Hukou is from non-rice farming provinces. The results are the same if we drop subjects whose Hukou province was different from the experimental province.

[18]There were 9 subjects who were not first year students. Including or excluding them does not affect our results and hence we include them in the analyses.

13.3 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 encounter everyday is very similar to the situation in the PGG: each farmer has the incentive to free-ride 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 faces 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.

Note that the contributor only gains 0.4 points for each point contributed to the group account. Therefore, contributing nothing always gives the highest material payoff regardless of the contribution of other group members. On the other hand, each point contributed to the group's account increases the payoff of the whole group by 1.6 points, 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 which is 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.[19] 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.[20] 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.

[19]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.

[20]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.[21]

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. The 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 levels of strategic consideration. Since the responder can reject an offer in the UG while she 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.

The stag hunt game is a two-player simultaneous move coordination game. The payoff matrix is presented in table 16. Subjects can choose between hunting a stag or a hare. Stag is harder to catch but more valuable. Both

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

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.

Table 16: Payoff matrix of the Stag Hunt Game

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

Subjects' risk attitudes were elicited using Holt and Laury type lotteries (Holt and Laury, 2002). They were informed that this task is not incentivized. We chose not to incentivize the risk aversion task in order to ensure more ballanced payments to the subjects. If this task was incentivized, some subjects would leave the experiment only with the show up fee. If that happened, there was the risk of students not showing up for our sessions. In order to avoid that, we opted to incentivize the other tasks, where we expected better payoffs. We also implemented 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 monkeys eat bananas.

Since the literature in psychology suggests that collective societies value group membership, we conducted 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 the Hebei province.[22] In the other sessions, subjects are not informed about this. We find that priming has

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

no effect on subject's behavior, we therefore pool the data from priming and no-priming sessions in the analyses.

13.4 Experimental Procedure

After all subjects arrived in the lab and prior to getting any instructions about the study, they were asked to sign a formal consent. Participants knew that each session consists of 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 of the non-incentivized lottery task and the triad task. Then, they played one shot DG, UG, Stag Hunt games. After they completed a post experimental questionnaire, they received their payment and were free to leave. No feedback was provided to the subjects regarding the outcome of the DG, UG and Stag Hunt games until the end of the session. Next, they proceeded with participating in the PGG no punishment condition, and the PGG punishment condition, in that order.

The reason we set the order of the games as previously described is that we wanted to avoid the outcome of one game affecting 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 condition, the 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 administered the punishment condition last. In sum, the order of the games aims to minimize the spillover 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 pe-

riods. The subjects' earnings were exchanged to Chinese Yuan at the rate: 1 points = 0.5 Yuan (about 8 US Cents).

The experiments were conducted between Oct 2015 and Jan 2016 in China. All the tasks are programmed using z-Tree (Fischbacher, 2007). There were a total of 524 subjects. 116 subjects in 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 students' 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.[23]

14 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 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.

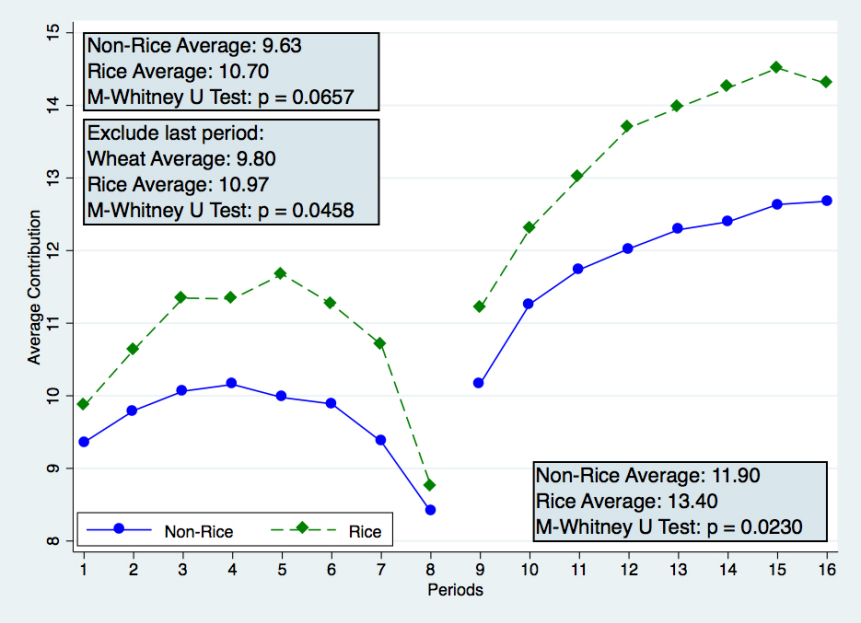
Result 12 *In the PGG without punishment, rice subject's contribution is higher than non-rice subject's contribution. However, the difference is not observed in the last period.*

Support for result 1 is presented in figure 13. Figure 13 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 indepen-

[23]See Appendix **??** 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.

dent observation: p = 0.0657).[24]

Figure 13: Average contribution in each period in PGG and PGG with Punishment



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

Results from Random Effects Panel Regressions are also consistent with the non-parametric tests. Columns 1 and 2 show the result for the no punishment condition. Note that the main variable of interest is Perc. Paddy Field (Municipality), which is the percentage of cultivated land devoted to rice paddy field at the *municipality* level. Perc. Paddy Field (Municipality) is a finer level of rice statistics than the percentage of rice paddy fields at the province level. Moreover, for all the regressions in which this finer level of variable are used, the results are the same if we use the rice dummy (province level) variable. According to column 1, a ten percentage point increase in the percent of cultivated land devoted to paddy field leads to a 1.2 point increase in contribution. This effect is weakly significant at 10% level.

[24]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 analysis. Note that the results do not change if we conduct Bootstrapped ttest on group level clusters.

The results are the same if we add a set of control variables in column 2. The set of control variables and their descriptions are presented in table 17.[25]

All of our results are robust to controlling for the municipality level GDP per capita. This is important because differences in economic development and market integration are strongly associated with the level of cooperation and trust Henrich et al. (2001, 2010). Also note that 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 cultural origins (Wen et al., 2004), and live in the same political institutions.[26] These evidence alongside with more robustness check exercises conducted in the Alternative Explanation section suggest that our findings are not merely a correlation.

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: $p = 0.0937$). Differences begin to emerge over time. The contribution of non-rice subjects increases modestly over interaction. In contrast, rice subjects manage to increase their contributions dramatically. However, rice subjects

[25]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 had 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 keep these subjects in the regressions without control variables.

[26]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 present biased time preferences. Bó et al. (2010) discover that people are more cooperative in social dilemma situations under democratic institutions.

Table 17: 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".

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: $p > 0.66$).[27]

The different contribution pattern between rice and non-rice is also confirmed in the random effects panel regression showed in column 1 and 2 of table 18. In column 1, both the Period and Period Squared are highly significant, suggesting that the contribution pattern of non-rice subjects exhibits an inverted-U shape. The interaction terms between the Rice dummy and the period terms are also significant, which implies that the rice subject's contribution pattern has more curvature. The results are the same if we add a set of control variables (column 2). Moreover, the contribution pattern between rice and non-rice is similar to the punishment condition (see columns 3 and 4).

The stronger end game effect of rice subjects compared to non-rice subjects in the no punishment condition suggests that rice farming does 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 contributions between rice and non-rice to be 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). Subjects' behavior in the PGG with punishment is summarized in result 2.

Result 13 *In PGG with punishment, rice subjects' contribution is significantly higher than non-rice subjects' contribution. The difference is already significant in the first period and becomes greater towards the end.*

[27]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: $p = 0.0458$).

Table 18: Random Effects Panel Regression about the different contribution patterns in rice and non-rice provinces

	No Punishment Condition		Punishment Condition	
	(1)	(2)	(3)	(4)
Rice	-0.279	-0.345	-3.817	-3.956
	(0.536)	(0.539)	(3.463)	(3.502)
Period	0.816***	0.816***	1.875***	1.875***
	(0.209)	(0.210)	(0.383)	(0.384)
Rice x Period	0.773**	0.773**	0.759	0.759
	(0.338)	(0.338)	(0.602)	(0.603)
Period Squared	-0.103***	-0.103***	-0.0623***	-0.0623***
	(0.0214)	(0.0214)	(0.0143)	(0.0143)
Rice x Period Squared	-0.0828**	-0.0828**	-0.0253	-0.0253
	(0.0341)	(0.0342)	(0.0231)	(0.0231)
From Rural (Dummy)		-0.275		-0.410
		(0.411)		(0.412)
GDP per cap. (Municipality)		-0.0316		0.0464
		(0.0596)		(0.0608)
Priming (Dummy)		-0.468		-0.158
		(0.542)		(0.567)
Holistic Thinking		1.328*		0.311
		(0.766)		(0.735)
Natural Science (Dummy)		0.653*		1.313***
		(0.382)		(0.392)
Collectivistic		0.488		0.676
		(0.443)		(0.466)
Individualistic		-0.629		-0.431
		(0.427)		(0.388)
Public Order (Belief)		-0.0336		-0.0448
		(0.0922)		(0.0796)
Trustworthy (Belief)		0.0777		0.123
		(0.0897)		(0.104)
Male		2.029***		2.159***
		(0.462)		(0.369)
Relative Income		-0.564**		-0.295
		(0.279)		(0.237)
Risk Attitude		0.0726		0.159
		(0.117)		(0.106)
Single Child (Dummy)		-0.364		-0.0905
		(0.478)		(0.409)
Constant	8.575***	8.029***	-1.524	-4.812
	(0.351)	(2.567)	(2.235)	(3.410)
Observations	4112	4112	4112	4112

Notes: Columns 1 and 2 are from the no punishment condition; columns 3 and 4 are the punishment condition. Note that there are 4112 observations, which implies 514 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. Another three subjects were dropped because the values of some control variables are missing, as explained in footnote 14. Clustered Standard errors (PGG group level) are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

The evidence for result 2 is presented in figure 13, in the part that refers to periods 9-16. The average 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.

Results from Random Effects Panel Regressions are also consistent with the non-parametric tests. Column 3 shows that a 10% increase in the percentage of cultivated land devoted to paddy field is associated with a 1.8 increase in contribution, and this relation is significant at the 5% level. This finding is robust to the inclusion of control variables (column 2).

One important feature of figure 13 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 14 *The presence of the punishment mechanism significantly increases the 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).[28] The results also suggest that punishment has a stronger effect for rice subjects. The difference in contribution between the no-punishment and punishment condition is larger for

[28]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.

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

In the following, we investigate subjects' 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 is 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.[29].

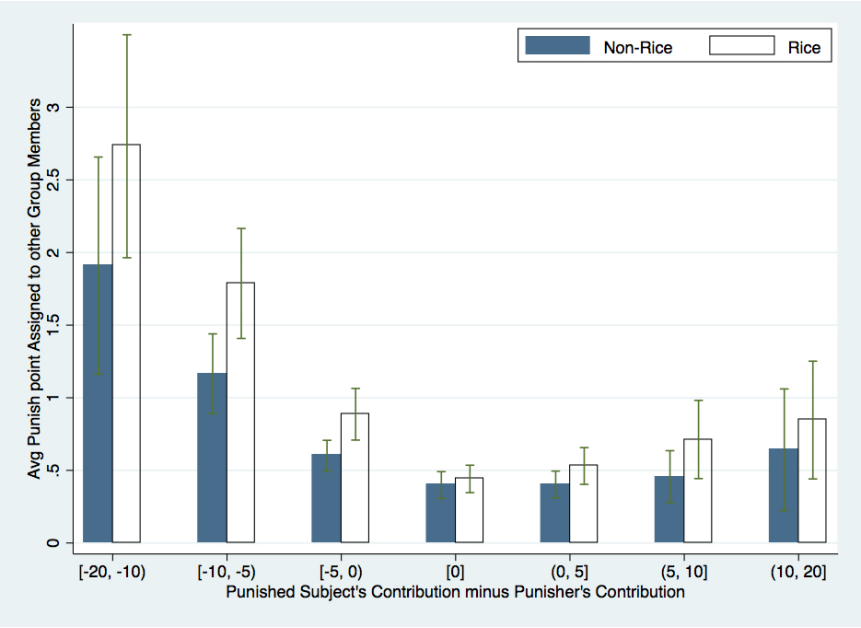
Result 15 *Rice subjects engage in significantly more pro-social punishment than non-rice subjects. There is no difference in antisocial punishment.*

Evidence for Result 15 is provided in figure 14. Figure 14 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

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

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) 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]).[30]

Figure 14: Punishment Behavior



Notes: This figure shows how subjects punish those who contribute less than, more than, 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.

The punishment pattern in figure 14 also suggests the necessity to separate

[30]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, $p = 0.08$).

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.

The results of tobit regressions regarding subject's punishment behavior are congruent with the non-parametric tests.[31] Column 1 indicates 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 suggest that Perc. of Paddy field is not associated with anti-social punishment ($P > 0.27$). The results are the same after including the set of control variables (column 2).

Result 16 *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 supporting Result 16 is presented in table 19. Columns 1-3 of table 19 show that rice subjects are more likely to punish free-riders in comparison to their non-rice counterparts.[32] And Columns 4-6 of table 19 shows that conditional on punishing, the intensity of punishment does not differ between rice and non-rice subjects. Note that the results are the same if we use the finer level of rice statistics instead of the rice dummy variable.

In column 1, the Rice dummy coefficient suggest that compared to their non-rice counterparts, subjects from rice areas are about 7% more likely to punish free-riders. The effect is weakly significant at the 10% level. The Anti-Social Punishment variable is significantly negative, but the interaction term between the Rice dummy and Anti-Social Punishment is not

[31]Note that following Herrmann et al. (2008), we classify anti-social punishment as the cases in which the punisher's contribution is strictly less than the punished subject's contribution. Therefore, cases when the punisher and the punished subject have the same contribution are not included in the regression. Results are the same if we include those observations and treat them as anti-social punishment.

[32]We use the linear probability model for the ease of presentation. The results are the same if we use the random effects panel probit model.

signifiant. This suggests that the likelihood of anti-social punishment is similar between rice and non-rice subjects.

In Column 2, after we control for the punisher's contribution, the punished subject's contribution, and the other two group member's average contribution, the effect becomes significant at 5%. 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 a punisher contributes 18 points in one period. In the first scenario, one group member contributes 15 points. In the second scenario, one group member contributes 8 points. Then the same punisher is more likely to punish the free-rider in the second scenario, simply because the difference in contribution is larger. To show that rice subjects are more likely to punish free-riders than non-rice subjects, we need to control for the punisher and punished subject's contribution. The reason to include the other two group member's average contribution is similar. The results are the same if we add the set of control variables (column 3).

Columns 4-6 of table 19 show that conditional on punishing, the intensity of punishment does not differ between rice and non-rice subjects.[33] 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-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 as well as other control variables (columns 5 and 6).

Result 17 *The percentage of cultivated land devoted to rice farming also*

[33]We used the 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.

Table 19: Regressions Regarding the Probability of Punishment and the Intensity of Punishment Conditional on Punishing

	Probabiliy of Punish			Intensity of Punishment Conditional on Punishing		
	(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)
Trustworthy (Belief)			-0.00920			0.0222
			(0.00810)			(0.0590)
Public Order (Belief)			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)
sigma						
Constant				1.865***	1.853***	1.754***
				(0.168)	(0.168)	(0.137)
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. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

predicts pro-social punishment within rice regions; there is no such relationship within non-rice regions.

The evidence supporting result 17 is presented in table 20. Columns 1 and 2 shows that among rice farming regions, a 10% increase in the percentage of cultivated land devoted to paddy fields is associated with about 0.7 additional punishment points assigned to free-riders, and this relation is significant at the 5% level. On the other hand, there is no such relationship among non-rice regions (columns 3 and 4).We did not find a positive relationship between the Percentage of paddy field and the level of contribution even among rice regions. Our interpretation of these findings is twofold. First, we believe this result to show that the level of contribution is influenced by a province wide norm. Second, the fact that subjects from rural areas in rice provinces punish more is indicative of the channel through which the norm operates. It stands to reason that those subjects will have more strongly ingrained the mechanism that brings about the norm and that is indeed manifested in their behavior in our games.

Result 18 *There is no difference between rice and non-rice in how subjects respond to punishment.*

Table 21 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 contributed less than the 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 subjects significantly increase their contribution after receiving pro-social punishment. In particular, they contribute

Table 20: Tobit regressions regarding the relationship between the percentage of paddy fields and punishment behavior *within* rice and non-rice 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 table shows how does percentage of paddy field predict punishment behavior within rice and non-rice regions. Columns 1-3 are data from rice regions and columns 4-6 are data from rice regions. Cluster Standard errors (PGG group level) are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

0.675 more points for each pro-social punishment 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 they 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 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 their response to anti-social punishment.

15 Alternative Explanations

15.1 Rice farming and Social Preferences

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. One possible explanation is that individuals from rice and non-rice regions develop different social preferences. For example, rice subjects might be less tolerant towards inequality (Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999). Since free-riders earn more than cooperators in the PGGs, inequality aversion can explain why rice subjects are more likely to punish free-riders. Another possible explanation is that rice subjects might put more weight on social welfare or efficiency concerns (Charness and Rabin, 2002). In our experimental setting, each contribution increases the total earning of the group by three points, while not contributing will only increase total welfare by one point. Therefore, rice subjects are more likely to punish free-riders because pro-social punishment has shown to be one of the most effective ways to foster cooperation (Fehr and Gächter, 2000; Masclet et al., 2003). Moreover, since punishing cooperators is detrimental to contribution (Herrmann et al., 2008), this might also explain why rice subjects are *not* more likely to conduct anti-social punishment. Lastly, rice and non-rice might differ in the level of their reciprocity concerns (Falk and Fischbacher, 2006; Rabin, 1993).

Table 21: 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)
Trustworthy (Belief)		0.0192		0.0805
		(0.0412)		(0.0531)
Public Order (Belief)		-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. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

In this subsection, using data from the dictator game, ultimatum game, stag hunt game, PGGs, as well as questionnaire measures, we show that our results are less likely to be driven by these alternative explanations. This evidence alongside with the findings articulated in the Results section provide further support for the case suggesting that rice cultivation does not modify people's cooperative preferences per se, instead, it teaches people the effectiveness of punishment in fostering cooperation in social dilemma situations.[34]

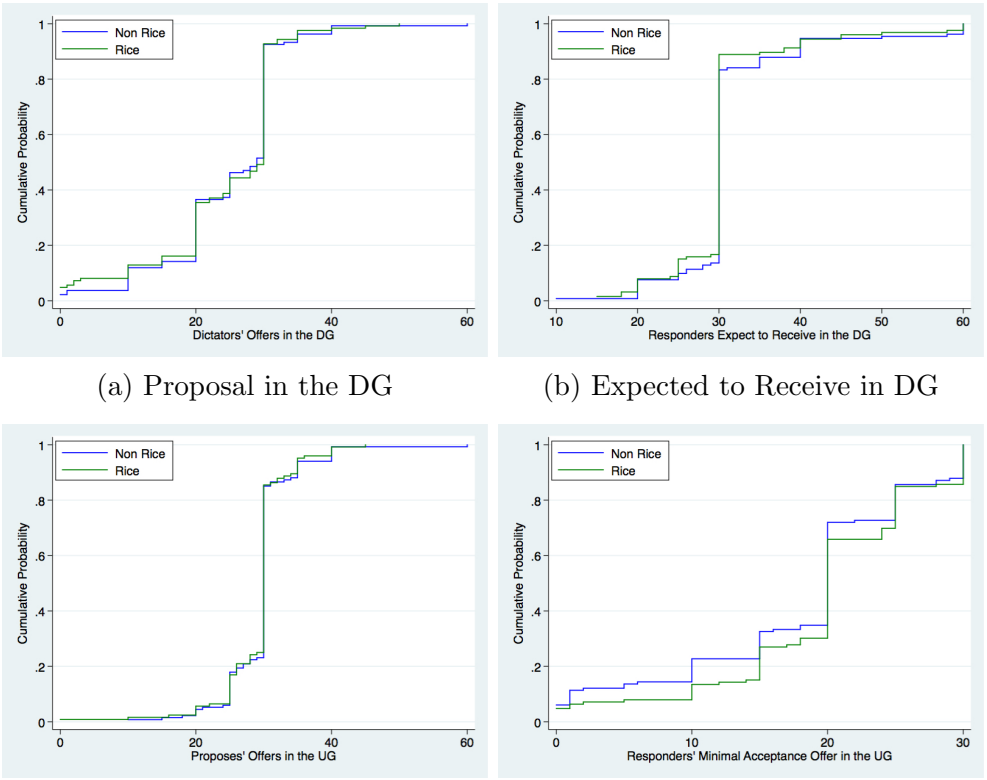
Results from the dictator game and ultimatum game, which are presented in figure 15, suggest that pure altruism and inequality aversion is not the main difference between rice and non-rice subjects. Panel (a) and panel (c) of figure 15 present the distribution of offers in the dictator game and ultimatum 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).

When measuring people's reciprocity concerns, 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 is not an important factor in explaining the results.

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

[34]It is not our intention to test for which social preferences model best fits the behavior of our subjects. Instead, we merely wish to point out that rice and non-rice subjects do not differ in these preferences, and hence the difference in behavior between them are mediated via different channels.

Figure 15: 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 expected 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.

(Mann whitney U test, each group as an independent observation: p = 0.07). Second, in the dictator game, we ask the responders the amount they expect to receive from the dictators. Since responder's expectation has no influence on dictator's behavior (recall that dictators do not know this procedure) as well as the outcome of the game, this essentially measures responder's belief on dictator's social preference. Responder's beliefs are similar 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: no-rice = 72.7%; rice = 72.4%; $p > 0.94$). Note that 'hare' in the stage hunt game is a risk dominant action, which means if one is sufficiently uncertain of the 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 concerns 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 the following exercises.

Following Croson (2007), we regress subject's contribution in period t on other group member's Min, Median, and Max contribution in the *same* period. This analysis is able to investigate subjects' reciprocity concerns because Croson (2007) find that subjects' beliefs about the amount other group members would contribute in the next period in the PGG is quite accurate.[35]

The results are shown in table 22. They illustrate that there is no difference in the level of reciprocity between rice and non-rice subjects. Columns 1 and 2 show the results for the no punishment condition. The other group member's Min, Median, and Max are all positively signifiant. However, none of the interaction terms is signifiant at any conventional levels. This is also true in the punishment condition (columns 3 and 4). These results

[35]Since we did not elicit subjects' beliefs, we cannot be sure that subjects in our sample can make accurate inference about others' behavior. Therefore, we also regress subject's contribution in period t on other group member's Min, Median, and Max contribution in the *previous* period $(t-1)$. The results are the same.

alongside with the finding that rice and non-rice subjects hold similar beliefs suggest that different levels of reciprocity concerns cannot account for the behavioral differences in the PGGs between rice and non-rice subjects.

Table 22: Random Effects Panel Regressions about reciprocity in the PGGs

	No Punishmet		Punishment	
	(1)	(2)	(3)	(4)
Rice	0.412	0.224	0.509	0.149
	(0.775)	(0.769)	(0.921)	(0.884)
Other Median	0.260***	0.258***	0.426***	0.418***
	(0.0694)	(0.0668)	(0.0657)	(0.0655)
Rice x Other Median	-0.0909	-0.0943	-0.0684	-0.0645
	(0.0902)	(0.0879)	(0.0844)	(0.0834)
Other Max	0.0737*	0.0662	0.00964	0.000837
	(0.0443)	(0.0436)	(0.0515)	(0.0504)
Rice x Other Max	0.0536	0.0603	0.0707	0.0791
	(0.0646)	(0.0638)	(0.0701)	(0.0685)
Other Min	0.0766*	0.0690*	0.113**	0.108**
	(0.0432)	(0.0414)	(0.0440)	(0.0453)
Rice x Other Min	0.0587	0.0648	0.00403	0.00961
	(0.0643)	(0.0621)	(0.0551)	(0.0553)
Period	0.666***	0.679***	0.722***	0.744***
	(0.114)	(0.115)	(0.195)	(0.201)
Period square	-0.0809***	-0.0825***	-0.0234***	-0.0241***
	(0.0125)	(0.0126)	(0.00726)	(0.00750)
From Rural (Dummy)		-0.186		-0.210
		(0.373)		(0.290)
GDP per cap. (Municipality)		-0.0251		0.0170
		(0.0439)		(0.0367)
Priming (Dummy)		-0.273		-0.119
		(0.315)		(0.258)
Holistic Thinking		0.871		0.233
		(0.626)		(0.528)
Natural Science (Dummy)		0.528*		0.871***
		(0.312)		(0.262)
Collectivistic		0.336		0.333
		(0.398)		(0.326)
Individualistic		-0.483		-0.282
		(0.374)		(0.305)
Public Order (Belief)		0.00581		-0.0144
		(0.0765)		(0.0577)
Trustworthy (Belief)		0.0397		0.0269
		(0.0763)		(0.0785)
Male		1.616***		1.633***
		(0.392)		(0.262)
Relative Income		-0.396		-0.351*
		(0.255)		(0.190)
Risk Attitude		0.0684		0.0799
		(0.0893)		(0.0730)
Single Child (Dummy)		-0.108		0.159
		(0.409)		(0.261)
Constant	4.768***	4.462*	0.270	-0.808
	(0.551)	(2.353)	(1.467)	(2.481)
Observations	4112	4112	4112	4112

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

15.2 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 born in a rice province but themselves have Hukou from a non-rice province or vise versa. Differently put, we drop subjects who have probably migrated into or out of rice farming regions.

This strategy can control for the self selection issue because prior to 1990, the Chinese government was enforcing laws that restricted immigration and travel within China.[36] Also note that China has had strict immigration controls 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 fathers were born before the time when the immigration law was lifted.[37] Consequently, the birth place of the 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. 8 subjects are dropped, and the results are not affected from excluding them.

15.3 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

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

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

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 cultural origins (Wen et al., 2004), and live in the same political institutions.[38] Moreover, we also controlled for a set of variables including municipality level GDP per capita, which suggest that our results are not driven by the difference in economic development and market integration (Henrich et al., 2001, 2010).

Secondly, 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 23 and table 24 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 the two non-rice provinces.

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

16 Complementary Evidence

In order to demonstrate the external validity of our experimental results we provide here evidence from two unrelated surveys and a natural field experiment. In the following we describe the measures we used and present evidence that is consistent with our interpretation of the experimental results.

[38]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 between future and present events leads to more present biased time preferences. Bó et al. (2010) discover that people are more cooperative in social dilemma situations under democratic institutions.

Table 23: 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 24: 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: The dependent variable is the amount of punishment points assigned to free-riders or cooperators. Cluster Standard errors (PGG group level) are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

16.1 A natural field experiment

Wikipedia is a global encyclopedia that relies on voluntary contributors to write its entries and provide content that anyone can enjoy. As such it is a prime example of a public good[39]. Despite some initial issues with blocked access from the Chinese authorities, Wikipedia has been largely accessible to mainland China since 2008[40]. In addition, according to data provided by Wikipedia, its simplified Chinese edition ranks in the bottom 10 of edits made by automated content editors (bots)[41]. Finally, Zhang and Zhu (2011) found that a temporary block on Wikipedia in mainland China in 2005 lead to a statistically significant reduction in contributions, indicating the existence of a community of Wikipedia users who reside in mainland China. Combining the aforementioned results with our current findings we hypothesise that users from rice areas are more likely to contribute to Wikipedia articles. Since the encyclopedia does not provide any data to identify the location of its contributors and IP addresses may not be very reliable given the availability of VPN workarounds in mainland China, we focus our attention to contributions made to the pages of cities. Contributors need to have both knowledge of the topic to the page of which they contribute and an interest in improving its presentation on Wikipedia. We assume that people who possess both qualities in sufficient levels to contribute to the pages of Chinese cities are more likely to hail from those areas, therefore offering us an indirect way to control for location. We used Wikipedia's own list of Chinese cities[42] to download data on the number of edits and the total size in bytes of the pages of those cities in the version of Wikipedia in simplified Chinese. Our sample contains 206 cities from mainland China. Of those cities 195 are eligible for our purposes, as they do not belong to regions with big minority populations. Table 25 presents the descriptive statistics of the sample of eligible cities. The variables of

[39]We were inspired to look into Wikipedia as a natural field experiment of a public good by Georganas and Li (2010)

[40]For more information see the documentation on Wikipedia itself: https://en.wikipedia.org/wiki/Censorship_of_Wikipedia#China

[41]For more information see: https://stats.wikimedia.org/EN/Sitemap.htm

[42]For a complete list of cities see: https://zh.wikipedia.org/wiki/%E4%B8%AD%E8%8F%AF%E4%BA%BA%E6%B0%91%E5%85%B1%E5%92%8C%E5%9C%8B%E5%9F%8E%E5%B8%82%E4%BA%BA%E5%8F%A3%E6%8E%92%E5%90%8D. Data accessed in March 2017.

interest are the Page size (in bytes) and the Total number of edits. The former is a measure of the intensive and the latter of the extensive margin of contributions to the public good. We use the population of the urban area of the city, the GDP per capita of the province in 2005 and the growth of the per capita GDP between 2005 and 2015 as additional controls. The inclusion of the GDP per capita from other years does not affect our results and in the interest of space we do not include these measures in our presentation. We use the binary measure for the classification of cities into rice and non-rice.

Table 25: Descriptive statistics of the Wikipedia sample

Statistic	N	Mean	St. Dev.	Min	Max
Non-rice Provinces					
Population (Urban area)	83	1,935,605.000	2,526,350.000	328,555	19,295,000
Page size (bytes)	83	34,062.450	36,605.520	24	211,385
Total number of edits	83	472.060	716.669	1	4,752
GDP_{2005}	83	16,142.880	5,622.360	7,477	45,444
g_{2005_2015}	83	2.293	0.485	1.343	3.811
Rice Provinces					
Population (Urban area)	112	2,251,961.000	2,829,462.000	155,540	22,265,426
Page size (bytes)	112	40,381.000	40,536.060	2,155	181,363
Total number of edits	112	595.929	763.794	40	4,794
GDP_{2005}	112	18,004.040	8,489.157	5,052	51,474
g_{2005_2015}	112	2.632	0.651	1.016	4.908

Result 19 *The pages of cities in rice provinces have more edits than cities from non-rice provinces, demonstrating a difference in the extensive margin of contribution to the public good.*

Support for this result is provided in table 26. Since the number of edits is a discrete and non-negative variable we use Poisson and Negative Binomial regressions used in the analysis of count data. The Rice dummy is positive and statistically significant at the 1% level in both the Poisson and the Negative Binomial regression (columns (2) and (3)). Column (1) also presents the results from an OLS regression where we applied a logarithmic transformation due to the discreteness of the dependent variable. Columns (4)-(6) show that the same results hold if we restrict the sample

to cities in the four provinces from our experiment. This shows both that our provinces are typical, in that respect, of rice and non-rice provinces and that our results extend beyond the lab.

Result 20 *The pages of cities in rice provinces have more bytes than cities from non-rice provinces, demonstrating a difference in the intensive margin of contribution to the public good.*

Support for this result is provided in table 27. For reasons similar to the ones about the edits we use Poisson and Negative Binomial regressions. The Rice dummy is positive and statistically significant at the 5% level in both the Poisson and the Negative Binomial regression (columns (2) and (3)). Column (1) also presents the results from an OLS regression where we applied a logarithmic transformation on the dependent variable. The rice variable is positive but not significant in this specification (p-value=0.12). Columns (4)-(6) show that the same results hold if we restrict the sample to cities in the four provinces from our experiment. Again this result is important for the external validity of our experimental findings both with respect to other provinces and within the provinces where we ran the experiment.

16.2 Volunteering

The World Values Survey (WVS) is a major survey in the social sciences investigating the attitudes of people all over the world. We used responses from the sixth wave of the survey, which was administered in 2012 in China and between 2010 and 2014 worldwide, to construct a measure of the volunteering propensity of responders in rice and non-rice regions. We chose the sixth wave of the survey because it has a large sample (N=2300), relatively evenly ballanced between rice and non-rice regions (1238 and 1062 responses respectively). Volunteering is akin to contributing to a public good; the volunteer is unlikely to be compensated for the time supplied unless a sufficiently large number of volunteers exists that she may find herself on the receiving end of others contributions. We hypothesise that

Table 26: Regressions about the total number of edits on the pages of cities in China on the Chinese Wikipedia

	Full sample			*Experimental provinces*		
	log(Edits)	Edits		log(Edits)	Edits	
	OLS	*Poisson*	*Negative binomial*	*OLS*	*Poisson*	*Negative binomial*
	(1)	(2)	(3)	(4)	(5)	(6)
Rice	0.297**	0.071***	0.307***	0.934***	0.690***	0.792***
	(0.120)	(0.007)	(0.101)	(0.315)	(0.018)	(0.212)
GDP_{2005}	−0.00001	−0.00000***	−0.00001	−0.0001	−0.00003***	−0.00004
	(0.00001)	(0.00000)	(0.00001)	(0.00003)	(0.00000)	(0.00002)
g_{2005_2015}	−0.126	0.047***	−0.170	−0.573	−0.099***	−0.344
	(0.127)	(0.007)	(0.107)	(0.442)	(0.028)	(0.298)
Population (log)	0.957***	1.008***	0.906***	1.024***	1.109***	0.924***
	(0.073)	(0.004)	(0.062)	(0.226)	(0.011)	(0.152)
Constant	−7.465***	−8.484***	−6.448***	−7.012**	−9.492***	−6.135***
	(1.040)	(0.055)	(0.877)	(3.329)	(0.178)	(2.243)
Observations	195	195	195	52	52	52

Notes: standard errors are reported in brackets. Because of the nature of count data the dependent variable was log transformed for the OLS regressions. Results from truncated Poisson and Negative binomial regressions are similar and available upon request. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

Table 27: Regressions about the size (in bytes) of pages of cities in China on the Chinese Wikipedia

	Full sample			*Experimental provinces*		
	log(Bytes)	Bytes		log(Bytes)	Bytes	
	OLS	*Poisson*	*Negative binomial*	*OLS*	*Poisson*	*Negative binomial*
	(1)	(2)	(3)	(4)	(5)	(6)
Rice	0.222	0.110***	0.245**	0.975**	0.567***	0.709***
	(0.142)	(0.001)	(0.115)	(0.380)	(0.002)	(0.246)
GDP_{2005}	−0.00003**	−0.00002***	−0.00003***	−0.0001**	−0.00003***	−0.0001**
	(0.00001)	(0.00000)	(0.00001)	(0.00004)	(0.00000)	(0.00003)
g_{2005_2015}	−0.183	−0.083***	−0.203*	−0.538	0.059***	−0.242
	(0.150)	(0.001)	(0.122)	(0.534)	(0.003)	(0.346)
Population (log)	0.754***	0.729***	0.719***	1.061***	1.180***	1.011***
	(0.087)	(0.0005)	(0.071)	(0.273)	(0.001)	(0.177)
Constant	0.228	0.443***	0.995	−2.650	−6.217***	−2.707
	(1.231)	(0.007)	(0.999)	(4.016)	(0.019)	(2.601)
Observations	195	195	195	52	52	52

Notes: standard errors are reported in brackets. Because of the nature of count data the dependent variable was log transformed for the OLS regressions. Results from truncated Poisson and Negative binomial regressions are similar and available upon request. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

people living in rice areas volunteer more. The WVS contains questions that allow us to test this hypothesis. More precisely, people were asked whether they are members (active or inactive) of eleven types of organizations, ranging from sport or recreational organizations, self help groups and environmental organizations to political parties, labor unions and professional associations.[43]. Using the answers to these questions we contructed two variables: the first is a binary variable that takes value 1 if the responder volunteers in at least one organization and zero otherwise. The second variable is equal to the sum of organizations to which the responder volunteers. In the same vein as our analyses above, the binary variable measures the extensive margin and the count variable measures the intensive margin of volunteering.

Descriptive statistics regarding the two variables are presented in table 28. Our control variables include the sex of the responder, age, education, the class and the income category to which she declared she belonged, whether her household saved money during the previous year, the number of children she has, her satisfaction with her financial condition, whether she lives in a city with a population greater than half a million and the GDP per capita of the province in 2011, the year before the survey was conducted. Due to reasons concerning privacy and anonymity we can only identify the province of the responders.

Result 21 *Respondents from rice provinces are more likely to volunteer.*

The evidence for this result is presented in columns (1) and (2) of table 29. Both the OLS and the logit regression show a strong ($p<0.01$) and positive relationship between the rice dummy and the probability of a respondent volunteering. We interpret this finding to mean that respondents from rice provinces show a difference in the extensive margin of volunteering.

Result 22 *Respondents from rice provinces volunteer to more organizations. However, this result is robust to different specifications only at the*

[43]We used the answers in variables V25 to V35 in the sixth wave of the World Values Survey

Table 28: Descriptive statistics of variables in WVS

Statistic	N	Mean	St. Dev.	Min	Max
Non-rice Provinces					
Volunteering Binary	1,062	0.218	0.413	0	1
Volunteering Count	1,062	0.459	1.273	0	11
Sex (Female=1)	1,062	0.524	0.500	0	1
Age	1,062	44.756	14.931	18	75
Education	1,062	5.111	2.196	1	9
Subjective Class	1,008	3.739	0.860	1	5
Subjective Income	924	4.378	2.001	1	10
Saved Money	985	1.734	0.822	1	4
Number of Children	1,062	1.443	1.025	0	7
Financial Satisfaction	1,059	6.475	2.028	1	10
CityPop>.5M	1,062	7.716	0.451	7	8
GDP 2011	1,062	38,419.410	13,881.250	19,595	81,658
Rice Provinces					
Volunteering Binary	1,238	0.281	0.450	0	1
Volunteering Count	1,238	0.587	1.319	0	11
Sex (Female=1)	1,238	0.498	0.500	0	1
Age	1,238	43.200	14.928	18	75
Education	1,238	5.532	2.488	1	9
Subjective Class	1,184	3.623	0.833	1	5
Subjective Income	1,131	4.447	1.723	1	9
Saved Money	1,171	1.679	0.801	1	4
Number of Children	1,238	1.463	1.096	0	7
Financial Satisfaction	1,186	5.985	1.947	1	10
CityPop>.5M	1,238	7.775	0.418	7	8
GDP 2011	1,238	41,013.540	17,991.950	16,413	82,560

10% level.

The evidence for this result is presented in columns (3) to (5) of table 29. Since we are dealing with count data we used the Poisson regression and also ran a Negative binomial and a Quasi-Poisson specification to correct for potential overdispersion. According to the Poisson regression the rice dummy is statistically significant at the 5% level. However, when we ran the alternative specifications it became significant only at the 10% level, suggesting that the relationship is not extremely robust.

16.3 A local public good

Finally, we utilized the China Family Panel Studies (CFPS), a large and nationally representative longitudinal survey of Chinese communities to provide evidence regarding the external validity of our experimental results. We used data of the CFPS from the 2010 and 2014 waves. Interviewers collecting data for CFPS visit the residences of the respondents. At the end of the questionnaire the interviewers are asked to grade the tidiness of the street of the interviees. The tidiness of the street is, arguably, a public good. The neighbors who take care not to litter the street and help maintain its tidiness volunteer their time and effort but they will only receive a reward at least equal to their contribution only if others contribute as well. Our dependent variable is the rating of the interviewers. Due to privacy concerns, we cannot identify the interviewer and we only have data aggregated at the province level. Our independent variables are the percentage of rice paddy fields in the province, the economic condition of the respondent, the number of households interviewed in the neighborhood, information on the architectural layout of the residence (house, appartment, etc) and three dummy variables about whether the interviewee live in an urban area, lives in a minority region and the wave of the survey.

Result 23 *The streets of respondents in rice areas are tidier.*

The evidence for this result is presented in columns (1) and (2) of table 30. The coefficient of the percentage of rice paddy fields cultivated in the

Table 29: Regressions regarding volunteering

	Volunteering Binary		Volunteering Count		
	OLS	*logistic*	*Poisson*	*Negative binomial*	*Quasi-Poisson*
	(1)	(2)	(3)	(4)	(5)
Rice	0.055***	0.315***	0.220***	0.206*	0.220*
	(0.020)	(0.115)	(0.069)	(0.117)	(0.113)
Sex	0.052***	0.291***	0.183***	0.179	0.183*
(Female=1)	(0.020)	(0.112)	(0.066)	(0.113)	(0.108)
Age	0.003***	0.016***	0.007**	0.009*	0.007
	(0.001)	(0.005)	(0.003)	(0.005)	(0.005)
Education	0.030***	0.163***	0.176***	0.167***	0.176***
	(0.005)	(0.029)	(0.018)	(0.030)	(0.029)
Subjective Class	−0.022	−0.126	−0.032	−0.047	−0.032
	(0.015)	(0.089)	(0.053)	(0.091)	(0.088)
Subjective Income	0.031***	0.178***	0.159***	0.155***	0.159***
	(0.007)	(0.042)	(0.025)	(0.042)	(0.041)
Saved Money	0.017	0.103	0.161***	0.144*	0.161**
	(0.013)	(0.077)	(0.044)	(0.077)	(0.072)
Number of Children	−0.032***	−0.194***	−0.170***	−0.191**	−0.170**
	(0.012)	(0.070)	(0.045)	(0.074)	(0.074)
Financial Satisfaction	−0.004	−0.021	0.015	0.019	0.015
	(0.006)	(0.032)	(0.020)	(0.033)	(0.033)
CityPop>.5M	0.054**	0.337**	0.435***	0.340**	0.435***
	(0.023)	(0.137)	(0.086)	(0.141)	(0.141)
GDP 2011	0.000	−0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	−0.515**	−5.741***	−6.492***	−5.647***	−6.492***
	(0.205)	(1.224)	(0.765)	(1.264)	(1.259)
Observations	1,873	1,873	1,873	1,873	1,873

Notes: standard errors are reported in brackets. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

province remains positive and statistically significant ($p<0.05$) even when controls are included. We interpret this as an indication that those living in rice provinces enjoy a local public good of higher quality. This is only possible if the contributions are also higher.

16.4 Discussion of complementary evidence

In this section we have provided evidence from three different sources regarding the external validity of our experimental results. The first piece of evidence concerns a natural field experiment about a global public good. We showed that the contributions to that public good were significantly higher in rice regions both on the intensive and the extensive margin. Wikipedia articles are a local public good, since everyone has potentially access to them. We believe that the use of data regarding the pages of cities is a good way to circumvent the problem of the widespread use of VPN services that render the use of the IP addresses of the anonymous editors potentially meaningless. Nevertheless, we must admit that in the ideal situation we would have much more personalised information. In order to address that concern, we turned to the WVS data. We believe that we have convincingly shown that volunteering in China, both on the intensive and the extensive margin, is higher in rice provinces. Finally, our last piece of evidence concerns the state of a public good: the tidiness of the streets of a neighborhood. Again, our results show that the streets in rice provinces are more tidy. Our results are show that whether we measure the inputs to a public good, as in our two pieces of evidence, or as an output, as in the last one, we find results that are fully consistent with the ones in our experiment. We believe that these findings, taken as a whole, provide strong support for the external validity of the experimental results.

17 Conclusion

This paper explores the origins of cultural differences in the level of cooperativeness. In particular, we show that the centuries old practice of rice farming in China leads to a cooperative social norm. This norm is intergen-

Table 30: Regressions regarding street tidiness

	Street tidiness	
	OLS	
	(1)	(2)
Perc. of Paddy Field	0.787***	0.187**
	(0.158)	(0.082)
Economic Condition		0.483***
		(0.030)
No. Households		0.00002
		(0.00002)
Minority Region (Dummy)		-0.375***
		(0.108)
Urban Area (Dummy)		0.097
		(0.059)
Architecture Layout		0.372***
		(0.030)
2014 Wave (Dummy)		-0.048
		(0.052)
Constant	4.412***	0.739***
	(0.079)	(0.101)
Observations	1254	1242

Notes: standard errors are reported in brackets. * $p < 0.1$, ** $p < 0.05$, ***$p < 0.01$

erationally transmittable. We go one step further from similar studies by providing evidence that the norm operates through the attitudes regarding punishment. To test the hypothesis, we travelled to four typical rice and non-rice farming provinces in China and recruited local university students who had no to 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. Importantly, we did not find any differences between rice and non-rice subjects in games that do not involve cooperation, such as the DG, UG, and Stag Hunt game. Additionaly, we found no differences in the intensity of punishment conditional on punishing between rice and non-rice subjects.

The results suggest that the cultivation of rice does not affect preferences for cooperation. Instead, it is more consistent with our findings that people in rice areas understand the public goods situation differently than their non-rice counterparts. There are three pieces of evidence supporting our conjecture.

First, the contributions of rice and non-rice subjects in the last round of the PGG without punishment are virtually identical. This is important because the strategic incentives to free ride are known to be strongest in the last round of a repeated game. Genuinely more cooperative subjects would contribute more in the last round, ignoring those incentives. Yet, what we observe is that their contributions do not differ. Second, the contributions of rice and non-rice subjects in the first round of the PGG without punishment are also very close and not statistically different from each other. In the first round, when nothing is known about the other group members, contributions are based on expectations. The fact that both subjects made very similar choices shows that their expectations do not differ[44]. Third, we measured social preferences and expectations regarding social preferences using standard dictator and ultimatum games. We find no differences in

[44]This conclusion is further reinforced by the fact that there are no differences within rice and non-rice subjects in the priming and no-priming conditions, where we emphasized the origins of the other participants.

those games. In addition to analyzing the experimental data we provide evidence from two surveys and a natural field experiment which show that the contributions of people in rice provinces are higher and the state of a local public good is improved. These results are consistent with our experimental findings and show that they are likely to be valid beyond the lab.

Furthermore, we show that subjects from rice areas are more likely to punish free-riders in the PGG's with punishment. However, there are no differences in the intensity of punishment conditional on punishing. We believe this shows a different attitude towards punishment. The reason that people from rice farming areas are more willing punish is probably the fact that punishment is both easier to enforce and more crucial in a rice farming environment. 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 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 to occur. The conditions are: positive individual gain from successful cooperation, member's action must be observable by others, and sanctions must be enforceable (Ray, 1998).

There are a number of open questions that are interesting for future research. For example, how is this norm affected by migration that brings in contact people from various backgrounds who carry different norms? Has the recent urbanization of China become a melting pot for norms and which ones dominate? Are market forces, through trade liberalization and industrialization, or homogenization efforts mandated by the state, such as the common curriculum in universities able to influence those norms and how? More broadly, to what other differences does this norm might lead? Do

people from rice and non-rice areas have different forms of social network? Do they behave differently in network games? Hsee and Weber (1999) find that Americans are more *risk averse* than Chinese and they proposed the "Cushion Hypothesis" arguing this is because they have different social networks – Chinese are more likely to receive help under financial pressure, therefore, they can afford to be more risk seeking. It would be interesting to see whether rice and non-rice subjects have different risk attitudes.

References

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.

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.

Ardagna, S. and F. Caselli (2014, October). The political economy of the greek debt crisis: A tale of two bailouts. *American Economic Journal: Macroeconomics 6*(4), 291–323.

Battaglini, M., R. B. Morton, and T. R. Palfrey (2010). The swing voter's curse in the laboratory. *The Review of Economic Studies 77*(1), 61–89.

Bhattacharya, S., J. Duffy, and S. Kim (2017, 3). Voting with endogenous information acquisition: Experimental evidence. *Games and Economic Behavior 102*, 316338.

Bhattacharya, S., J. Duffy, and S.-T. Kim (2014). Compulsory versus voluntary voting: An experimental study. *Games and Economic Behavior 84*(C), 111–131.

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.

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

Bosman, R. and F. van Winden (2002). Emotional hazard in a power-to-take experiment. *The Economic Journal 112*(476), 147–169.

Bouton, L., A. Llorente-Saguer, and F. Malherbe (2017). Unanimous rules in the laboratory. *Games and Economic Behavior 102*, 179 – 198.

Branas-Garza, P., P. Kujal, and B. Lenkei (2015). Cognitive Reflection Test: Whom, how, when.

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

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).

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.

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

Cueva, C., I. Iturbe-Ormaetxe, E. Mata-Prez, G. Ponti, M. Sartarelli, H. Yu, and V. Zhukova (2016). Cognitive (ir)reflection: New experimental evidence. *Journal of Behavioral and Experimental Economics 64*(Supplement C), 81 – 93. Cognitive Abilities and Economic Behavior.

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.

Dimmock, S. G., R. Kouwenberg, and P. P. Wakker (2016). Ambiguity attitudes in a large representative sample. *Management Science 62*(5), 1363–1380.

Elbittar, A., A. Gomberg, and T. R. Palfrey (forthcoming). Ignorance and bias in collective decisions. *Journal of Economic Behavior and Organization*.

Engelmann, D. and M. Strobel (2004, September). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review 94*(4), 857–869.

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.

Feddersen, T. J. and W. Pesendorfer (1996). The swing voter's curse. *American Economic Review 86*(3), 408–24.

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.

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.

Frederick, S. (2005, December). Cognitive reflection and decision making. *Journal of Economic Perspectives 19*(4), 25–42.

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.

Gaechter, S., E. J. Johnson, and A. Herrmann (2007). Individual-level loss aversion in riskless and risky choices. *Discussion Paper* (2961).

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

Georganas, S., P. J. Healy, and R. A. Weber (2015). On the persistence of strategic sophistication. *Journal of Economic Theory 159*(Part A), 369 – 400.

Georganas, S. and N. Li (2010). The altruism of nations: Measuring social capital in the field. *Mimeo*.

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.

Goeree, J. K. and L. Yariv (2011, May). An experimental study of collective deliberation. *Econometrica 79*, 893–921.

Greiner, B. (2004, June). The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics. Working Paper Series in Economics 10, University of Cologne, Department of Economics.

Grosser, J. and M. Seebauer (2016, 05). The curse of uninformed voting: An experimental study. *97*.

Guarnaschelli, S., R. D. McKelvey, and T. R. Palfrey (2000). An experimental study of jury decision rules. *American Political Science Review 94*(2), 407423.

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

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.

Herrera, H., A. Llorente-Saguer, and J. C. McMurray (2016). Information aggregation and turnout in proportional representation: A laboratory experiment. *Unpublished manuscript*.

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.

Hsee, C. and E. U. Weber (1999). Cross-national differences in risk preference and lay predictions. *Journal of Behavioral Decision Making 12*(2), 165–179.

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.

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

Johnson, E. J., C. Camerer, S. Sen, and T. Rymon (2002). Detecting failures of backward induction: Monitoring information search in sequential bargaining. *Journal of Economic Theory 104*(1), 16 – 47.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

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.

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.

Levitt, S. D., J. A. List, and S. E. Sadoff (2011, April). Checkmate: Exploring backward induction among chess players. *American Economic Review 101*(2), 975–90.

Ludbrook, J. (2008). Analysis of 2 2 tables of frequencies: matching test to experimental design. *International Journal of Epidemiology 37*(6), 1430–1435.

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.

Matsusaka, J. G. (1995). Explaining voter turnout patterns: An information theory. *Public Choice 84*(1/2), 91–117.

Mattozzi, A. and M. Y. Nakaguma (2016). Public versus secret voting in committees. *Unpublished manuscript*.

Mengel, F. and J. Rivas (2017). Common value elections with private information and informative priors: Theory and experiments. *Games and Economic Behavior 104*(C), 190–221.

Morton, R. B., M. Piovesan, and J.-R. Tyran (2013). The dark side of the vote: Biased voters, social information, and information aggregation through majority voting. (SP II 2013-209).

Morton, R. B. and J.-R. Tyran (2011, June). Let the experts decide? Asymmetric information, abstention, and coordination in standing committees. *Games and Economic Behavior 72*(2), 485–509.

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.

Oliveros, S. (2013, May). Abstention, ideology and information acquisition. *Jounral of Economic Theory 148*(3), 871–902.

Olivier, B. and L. Daniel (2013). Growth forecast errors and fiscal multipliers. *IMF Working Paper* (13/1).

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.

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

Soo Hong, C., R. P. Ebstein, and L. Yunfeng (2015). The rice culture theory of cooperative behavior: Evidence from incentivized decision making tasks in china. Working paper, National University of Singapore.

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.

Toplak, M. E., R. F. West, and K. E. Stanovich (2014). Assessing miserly information processing: An expansion of the cognitive reflection test. *Thinking & Reasoning 20*(2), 147–168.

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.

Zhang, X. M. and F. Zhu (2011, June). Group size and incentives to contribute: A natural experiment at chinese wikipedia. *American Economic Review 101*(4), 1601–15.

References

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.

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.

Ardagna, S. and F. Caselli (2014, October). The political economy of the greek debt crisis: A tale of two bailouts. *American Economic Journal: Macroeconomics 6*(4), 291–323.

Battaglini, M., R. B. Morton, and T. R. Palfrey (2010). The swing voter's curse in the laboratory. *The Review of Economic Studies 77*(1), 61–89.

Bhattacharya, S., J. Duffy, and S. Kim (2017, 3). Voting with endogenous information acquisition: Experimental evidence. *Games and Economic Behavior 102*, 316338.

Bhattacharya, S., J. Duffy, and S.-T. Kim (2014). Compulsory versus voluntary voting: An experimental study. *Games and Economic Behavior 84*(C), 111–131.

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.

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

Bosman, R. and F. van Winden (2002). Emotional hazard in a power-to-take experiment. *The Economic Journal 112*(476), 147–169.

Bouton, L., A. Llorente-Saguer, and F. Malherbe (2017). Unanimous rules in the laboratory. *Games and Economic Behavior 102*, 179 – 198.

Branas-Garza, P., P. Kujal, and B. Lenkei (2015). Cognitive Reflection Test: Whom, how, when.

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

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).

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.

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

Cueva, C., I. Iturbe-Ormaetxe, E. Mata-Prez, G. Ponti, M. Sartarelli, H. Yu, and V. Zhukova (2016). Cognitive (ir)reflection: New experimental evidence. *Journal of Behavioral and Experimental Economics 64*(Supplement C), 81 – 93. Cognitive Abilities and Economic Behavior.

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.

Dimmock, S. G., R. Kouwenberg, and P. P. Wakker (2016). Ambiguity attitudes in a large representative sample. *Management Science 62*(5), 1363–1380.

Elbittar, A., A. Gomberg, and T. R. Palfrey (forthcoming). Ignorance and bias in collective decisions. *Journal of Economic Behavior and Organization*.

Engelmann, D. and M. Strobel (2004, September). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review 94*(4), 857–869.

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.

Feddersen, T. J. and W. Pesendorfer (1996). The swing voter's curse. *American Economic Review 86*(3), 408–24.

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.

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.

Frederick, S. (2005, December). Cognitive reflection and decision making. *Journal of Economic Perspectives 19*(4), 25–42.

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.

Gaechter, S., E. J. Johnson, and A. Herrmann (2007). Individual-level loss aversion in riskless and risky choices. *Discussion Paper* (2961).

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

Georganas, S., P. J. Healy, and R. A. Weber (2015). On the persistence of strategic sophistication. *Journal of Economic Theory 159*(Part A), 369 – 400.

Georganas, S. and N. Li (2010). The altruism of nations: Measuring social capital in the field. *Mimeo*.

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.

Goeree, J. K. and L. Yariv (2011, May). An experimental study of collective deliberation. *Econometrica 79*, 893–921.

Greiner, B. (2004, June). The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics. Working Paper Series in Economics 10, University of Cologne, Department of Economics.

Grosser, J. and M. Seebauer (2016, 05). The curse of uninformed voting: An experimental study. *97*.

Guarnaschelli, S., R. D. McKelvey, and T. R. Palfrey (2000). An experimental study of jury decision rules. *American Political Science Review 94*(2), 407423.

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

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.

Herrera, H., A. Llorente-Saguer, and J. C. McMurray (2016). Information aggregation and turnout in proportional representation: A laboratory experiment. *Unpublished manuscript*.

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.

Hsee, C. and E. U. Weber (1999). Cross-national differences in risk preference and lay predictions. *Journal of Behavioral Decision Making 12*(2), 165–179.

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.

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

Johnson, E. J., C. Camerer, S. Sen, and T. Rymon (2002). Detecting failures of backward induction: Monitoring information search in sequential bargaining. *Journal of Economic Theory 104*(1), 16 – 47.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

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.

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.

Levitt, S. D., J. A. List, and S. E. Sadoff (2011, April). Checkmate: Exploring backward induction among chess players. *American Economic Review 101*(2), 975–90.

Ludbrook, J. (2008). Analysis of 2 2 tables of frequencies: matching test to experimental design. *International Journal of Epidemiology 37*(6), 1430–1435.

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.

Matsusaka, J. G. (1995). Explaining voter turnout patterns: An information theory. *Public Choice 84*(1/2), 91–117.

Mattozzi, A. and M. Y. Nakaguma (2016). Public versus secret voting in committees. *Unpublished manuscript*.

Mengel, F. and J. Rivas (2017). Common value elections with private information and informative priors: Theory and experiments. *Games and Economic Behavior 104*(C), 190–221.

Morton, R. B., M. Piovesan, and J.-R. Tyran (2013). The dark side of the vote: Biased voters, social information, and information aggregation through majority voting. (SP II 2013-209).

Morton, R. B. and J.-R. Tyran (2011, June). Let the experts decide? Asymmetric information, abstention, and coordination in standing committees. *Games and Economic Behavior 72*(2), 485–509.

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.

Oliveros, S. (2013, May). Abstention, ideology and information acquisition. *Jounral of Economic Theory 148*(3), 871–902.

Olivier, B. and L. Daniel (2013). Growth forecast errors and fiscal multipliers. *IMF Working Paper* (13/1).

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.

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

Soo Hong, C., R. P. Ebstein, and L. Yunfeng (2015). The rice culture theory of cooperative behavior: Evidence from incentivized decision making tasks in china. Working paper, National University of Singapore.

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.

Toplak, M. E., R. F. West, and K. E. Stanovich (2014). Assessing miserly information processing: An expansion of the cognitive reflection test. *Thinking & Reasoning 20*(2), 147–168.

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.

Zhang, X. M. and F. Zhu (2011, June). Group size and incentives to contribute: A natural experiment at chinese wikipedia. *American Economic Review 101*(4), 1601–15.

A Experimental Instructions to Chapter 1

Instructions

General Instructions

Welcome to the study!

If you have a question at any point, please raise your hand and wait for one of us to come over. We ask that you turn off your mobile phone and any other electronic devices. **Communication of any kind with other participants is not allowed**.

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. **Please read these instructions carefully**; in addition to the show up fee of £4, the cash payment you will receive at the end of the study depends on your and the other participants' decisions. So it is important that you read the instructions carefully. Your payment is calculated as follows:

Your payment = £4 (show-up fee) + earnings from 1st study + earnings from 2nd study

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

100 points = 40 pence.

The First Study

Summary

In every round the computer will choose a winning colour. This will be either green or yellow. In every round you will also be allocated into groups of 3. Your task will be to guess, as a group, which is the winning colour. **You only earn points if your group's guess matches the winning colour.** In every round you will also be assigned a green and a yellow value. If your group's guess matches the winning colour those values are used to determine how many point you will earn in that round.

Your group decides via majority rule. Before picking one option you can buy a hint regarding the winning colour. The accuracy of the hint is 80%. That is the hint will match the winning colour 4 out of 5 times. 1 out of 5 times the hint will not match the winning colour. The cost of the hint is 10 points.

Your earnings at the end of each round are as follows:

- If your group's guess and the winning colour are both green: earnings per round = green value - cost for hint (0 or 10 points)
- If your group's guess and the winning colour are both yellow: earnings per round = yellow value - cost for hint (0 or 10 points)
- If your group's guess and the winning colour differ: earnings per round = - cost for hint (0 or 10 points)

Winning Colour

At the beginning of each round the computer will randomly choose one of the two colours, yellow or green, to be the winning colour. **Both colours have a 50-50 chance of being the winning colour**.

Yellow and Green Value

At the beginning of every round you will also be assigned two values: a yellow value and a green value. **Your values are two numbers between 0 and 100.** Each number between 0 and 100 has an equal chance of being chosen. Just like you, the other members of your group are randomly assigned a green and a yellow value. **Their values need not be the same as yours.** You will not be informed at any point about the exact green and yellow values of the other group members. Your values and the other members' values are not in any way related to the computer's choice.

The yellow value is used to determine the amount of points you will earn if your group's guess and the winning colour are both yellow. The green value is used to determine the amount of points you will earn if your group's guess and the winning colour are both green.

Your Picks

In every round you will be asked to make a guess as a group. Each member of the group will be asked to pick one out of three options; pick yellow, pick green and pick Blank. **Blank picks are neutral and do not count in favour of either colour.** You will not be informed at

any point which option the other group members picked.

The group guess will be determined using a simple majority rule: **The colour (green or yellow) that was picked by the most members of the group will be the choice of your group.** In case of a tie, the computer will flip a coin to determine your group's guess.

The Hint

Before making your pick and after learning your green and yellow value for that round you will have the opportunity to **buy a hint.** In order to get the hint you will have to pay 10 points. If you buy the hint you will be shown a statement indicating which is the winning colour in that round. **The accuracy of the hint is 80%. That is the hint will match the winning colour 4 out of 5 times. 1 out of 5 times the hint will not match the winning colour.**

The hint is personal. That is you are the only one who knows your hint. The other members of your group will also have the opportunity to buy a hint. However you will not know at any point whether they did buy a hint and what that was.

Number of Rounds and Groups

There will be 30 rounds in this study. At the beginning of each round you will be allocated to a group of 3. In addition to yourself that group consists of 1 more human player and one **preprogrammed player**. After each round you will be allocated to a new group. The other human player in this new group need not be the same as the one in the previous round.

You will not know who the other participants in your group are at any point. In each round you will only deal with decisions in your group.

The Preprogrammed Player

Every time you play there is one preprogrammed player in your group. The preprogrammed player **never** buys a hint. In each round, **she picks either yellow or green with equal probability**. That is, in each round there is a 50% chance that she picks yellow and a 50% chance that she picks green. **The preprogrammed player's pick does not affect the round's winning colour nor does the winning colour affect the preprogrammed player's choice.**

Example 1

Your yellow value is 70. Your green value is 40. Your group's guess is yellow. The winning colour is yellow. Your pick was blank and you did not buy the hint. Your earnings in this round are:

yellow value - cost for hint = 70 - 0 = 70 points

Example 2

Your yellow value is 70. Your green value is 40. Your group's guess is green. The winning colour is yellow. Your pick was green and you bought the hint that cost you 10 points. Your earnings in this round are:

No match - cost for hint = 0 - 10 = -10 points

Example 3

Your yellow value is 70. Your green value is 40. Your group's guess is yellow. The winning colour is yellow. Your pick was yellow and you bought the hint that cost you 10 points. Your earnings in this round are:

yellow value - cost for hint = 70 - 10 = 60 points

Example 4

Your yellow value is 5. Your green value is 40. Your group's guess is yellow. The winning colour is yellow. Your pick was yellow and you bought the hint that cost you 10 points. Your earnings in this round are:

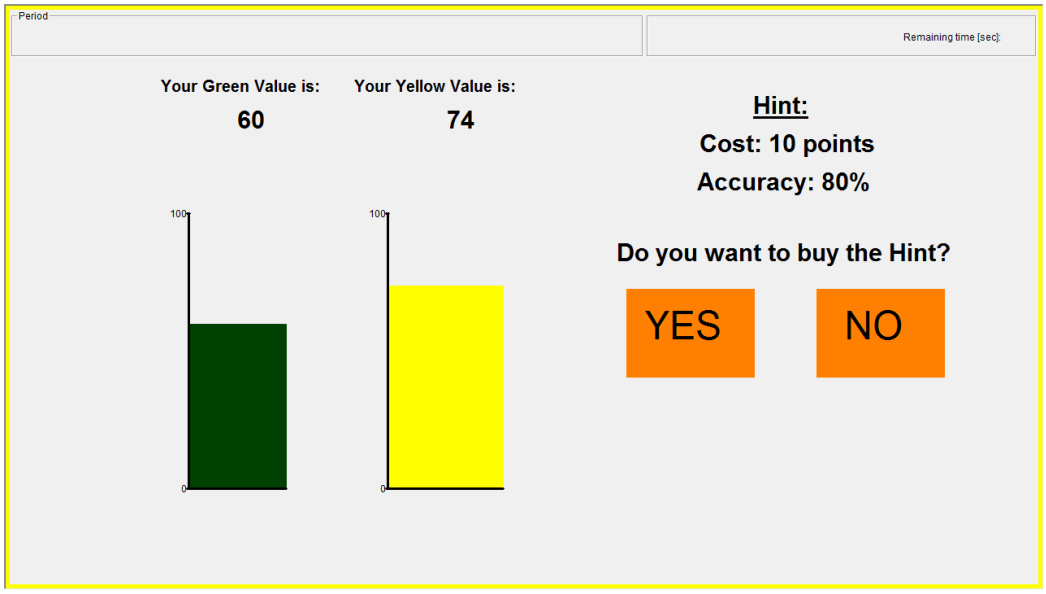
yellow value - cost for hint = 5 - 10 = -5 points

Details of the Procedure

Stage 1

In the picture below you can see a screenshot from stage 1. On the left side of the screen your green and yellow value for that round are displayed.

On the right side of the screen you can buy the hint. It costs 10 points and its accuracy is 80%. You can buy the hint by clicking on YES. If you don't want to buy the hint you can click on NO. **Please decide carefully. Once you click on a button you cannot change your decision.**



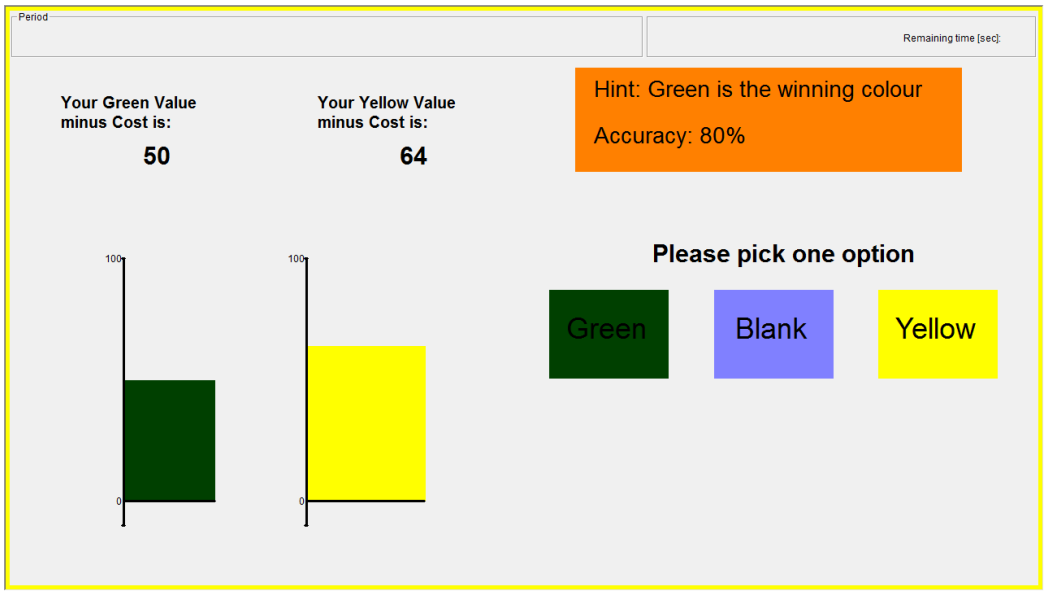
Stage 2

In the picture below you can see a screenshot from stage 2. On the left side of the screen you are informed about your green and yellow value minus the cost you paid for the hint in stage 1.

On the top right corner in the orange box you can see the hint, if you bought one, and its accuracy. The hint is a message saying:

yellow is the winning colour
or
green is the winning colour

On the middle right of the screen there are your three options: green, blank and yellow. You can pick one by clicking on the corresponding box. **Be advised: Once you click on a button you cannot change your pick.**

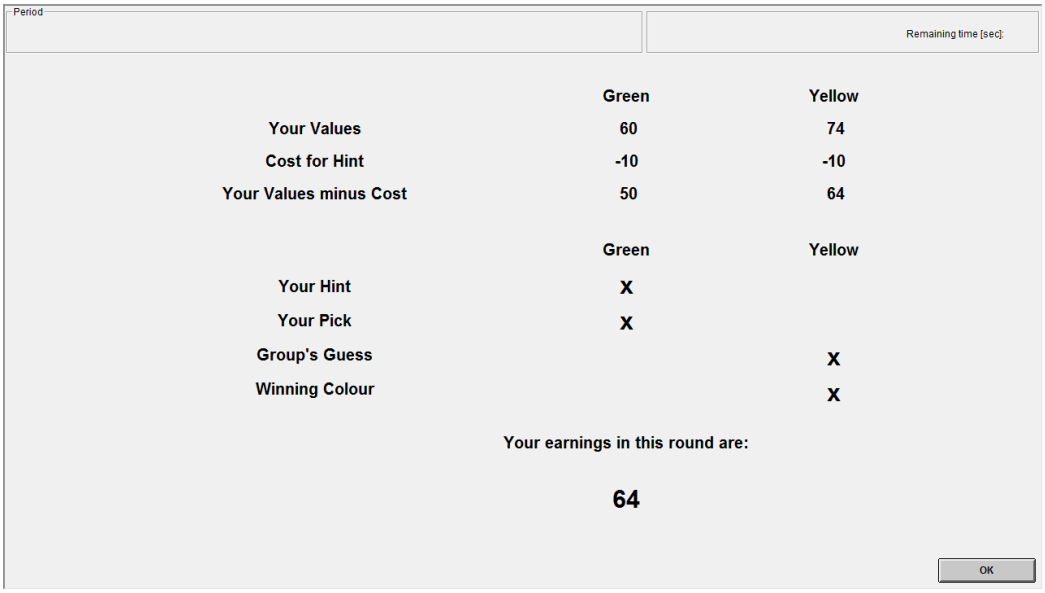


Stage 3

In the picture below you can see a screenshot from stage 3. On the top of this screen you are reminded of your yellow and green value in this round as well as the cost you paid for the hint.

On the middle you are reminded of what was the suggestion of your hint, if you bought one, and what was your pick. You are also told what was the guess of your group and the winning colour.

On the bottom of your screen you are told how may points you earned in this round.



Control Questions

1

Your yellow value is 60. Your green value is 30. What is the yellow value of the other human player in your group?

1. 60
2. 30
3. A number between 0 and 100
4. 0

2

Your yellow value is 60. Your green value is 30. You did not buy the hint. What is the winning colour?

1. Yellow
2. Green
3. Either Yellow or Green with a 50-50 chance

3

Your yellow value is 58. Your green value is 45. Assume that you picked yellow, the other human player picked blank and the winning colour is green. What option did the preprogrammed player pick?

1. Yellow
2. Green
3. Either Yellow or Green

4

Your yellow value is 60. Your green value is 30. You bought the hint and paid 10 points. You picked green and your group's guess is yellow. The winning colour is yellow. How many points did you earn in this round?

Answer:_____

5

Your yellow value is 20. Your green value is 90. You bought the hint and paid 10 points. You picked green and your group's guess is green. The winning colour is yellow. How many points did you earn in this round?

Answer:_____

6

Your yellow value is 36. Your green value is 44. You did not buy the hint. You picked Blank and your group's guess is green. The winning colour is green. How many points did you earn in this round?

Answer:_____

7

Your yellow value is 65. Your green value is 39. You bought the hint and paid 10 points. You picked Blank and your group's guess is yellow. The winning colour is green. How many points did you earn in this round?

Answer:_____

8

Your yellow value is 73. Your green value is 39. You did not buy the hint. You picked green and your group's guess is green. The winning colour is green. How many points did you earn in this round?

Answer:_____

Instructions for Second Study

The second study is very similar to the first study.

The only change is that **the preprogrammed player will always pick yellow.**

Everything else remains the same as in the previous rounds.

Again: **100 points = 40 pence**.

Control Question

1

Your yellow value is 60. Your green value is 45. You bought the hint and paid 10 points. Assume that you picked yellow, the other human player picked green and the winning colour is green.

What option did the preprogrammed player pick?

Answer:_____

What is your group's guess?

Answer:_____

What are your earnings in this round?

Answer:_____

Experimental Instructions to Chapter 3

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

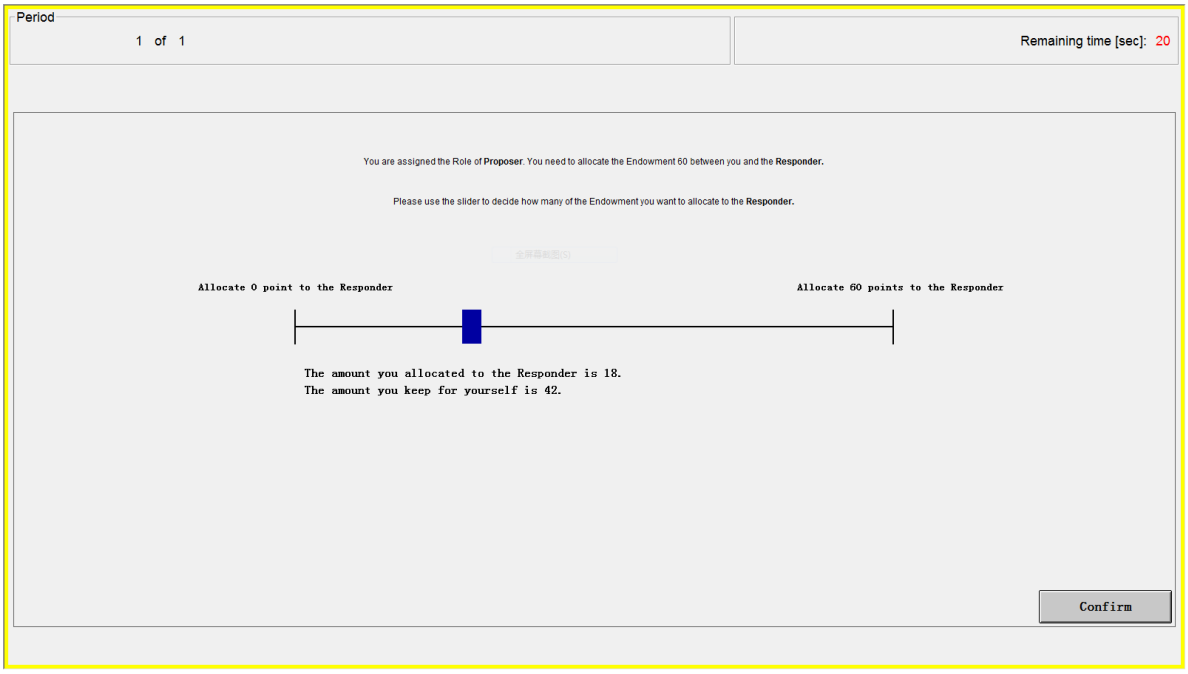
Please read the instructions carefully, because your earnings in each study depends on how well you understand the instructions.

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.

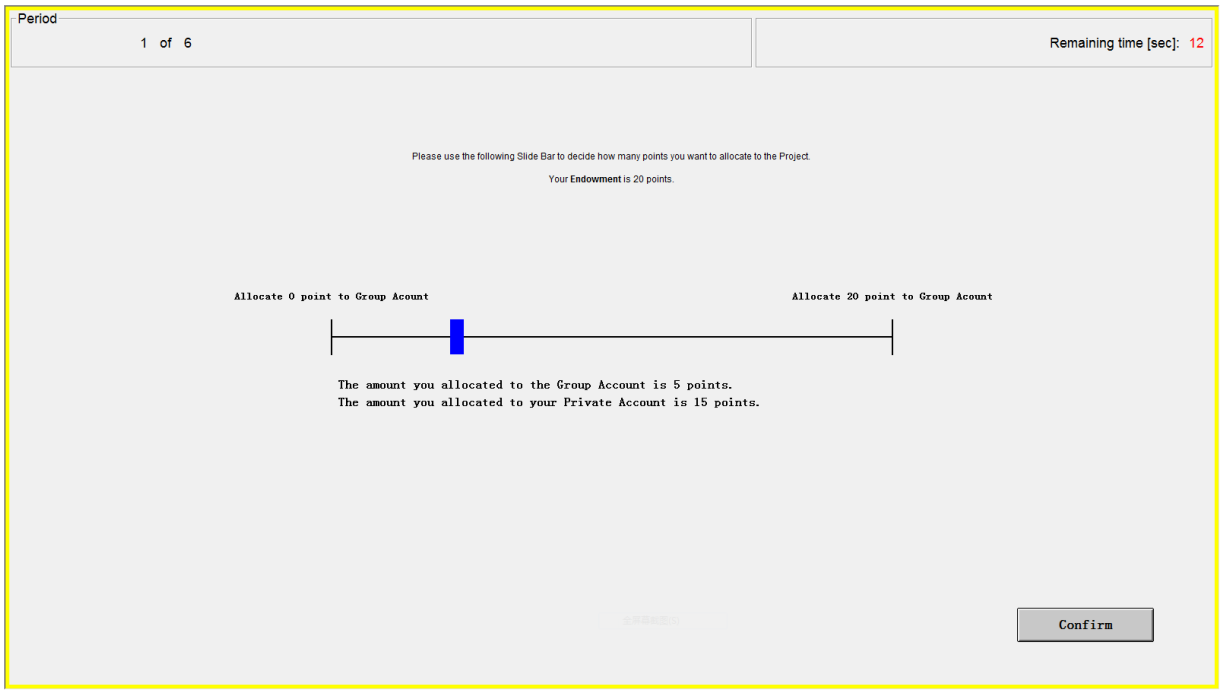
What was your choice?.....

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 =

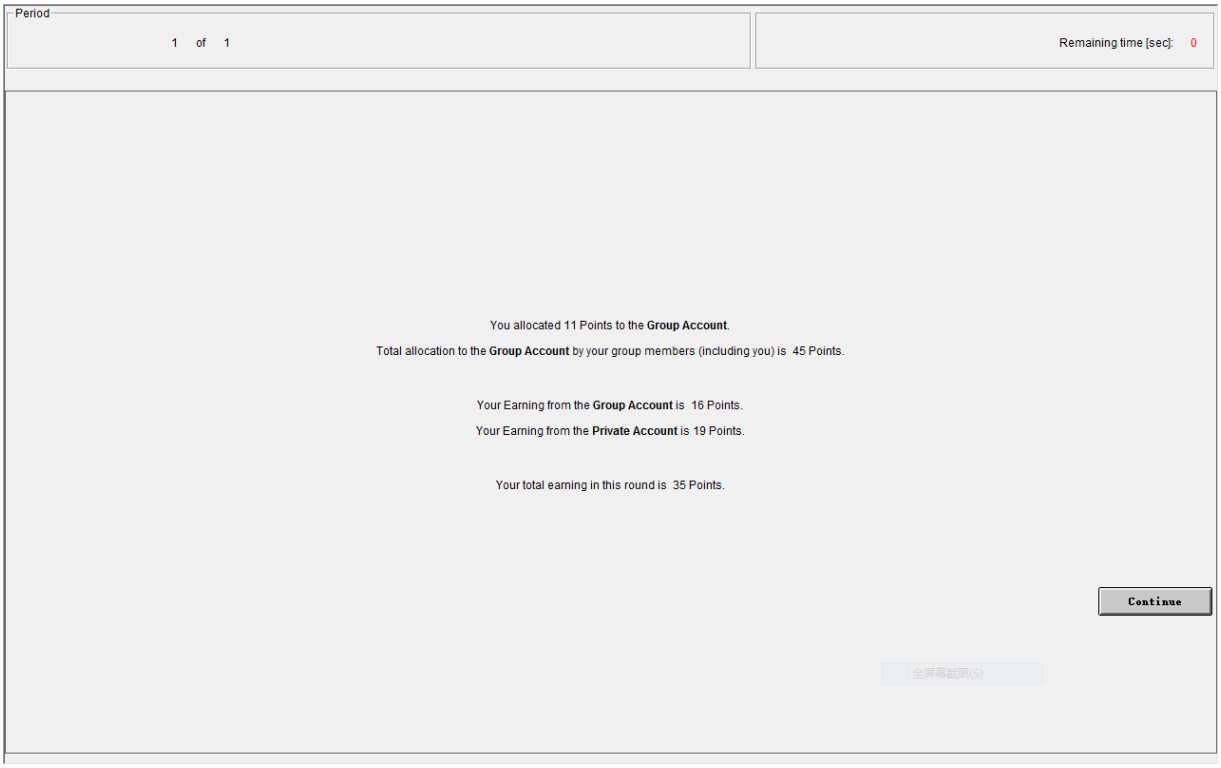
Earnings from Individual Account + Earnings from the Group Account =

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

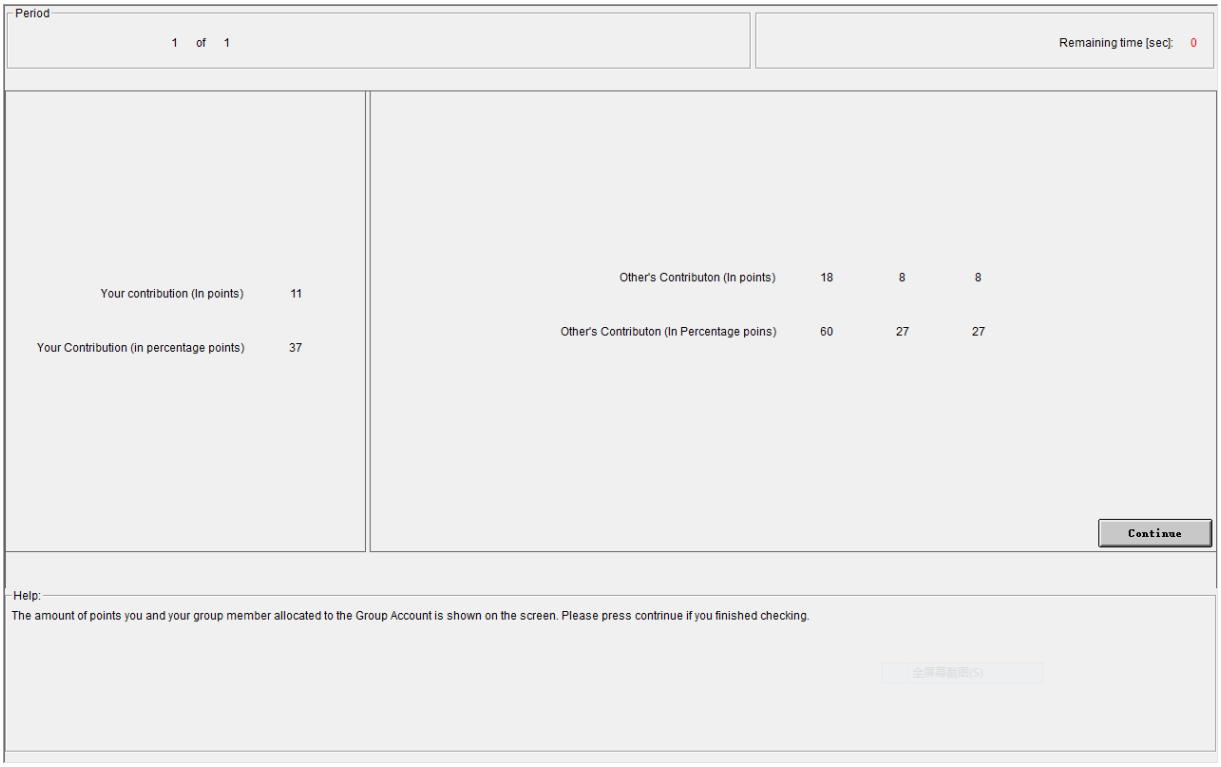
After all your group members have made their decision, your allocation and the sum of all allocations in your group are reported on the Result Screen as shown below. To aid you in your calculation, your earnings from your individual account and your earnings from the group account are both presented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

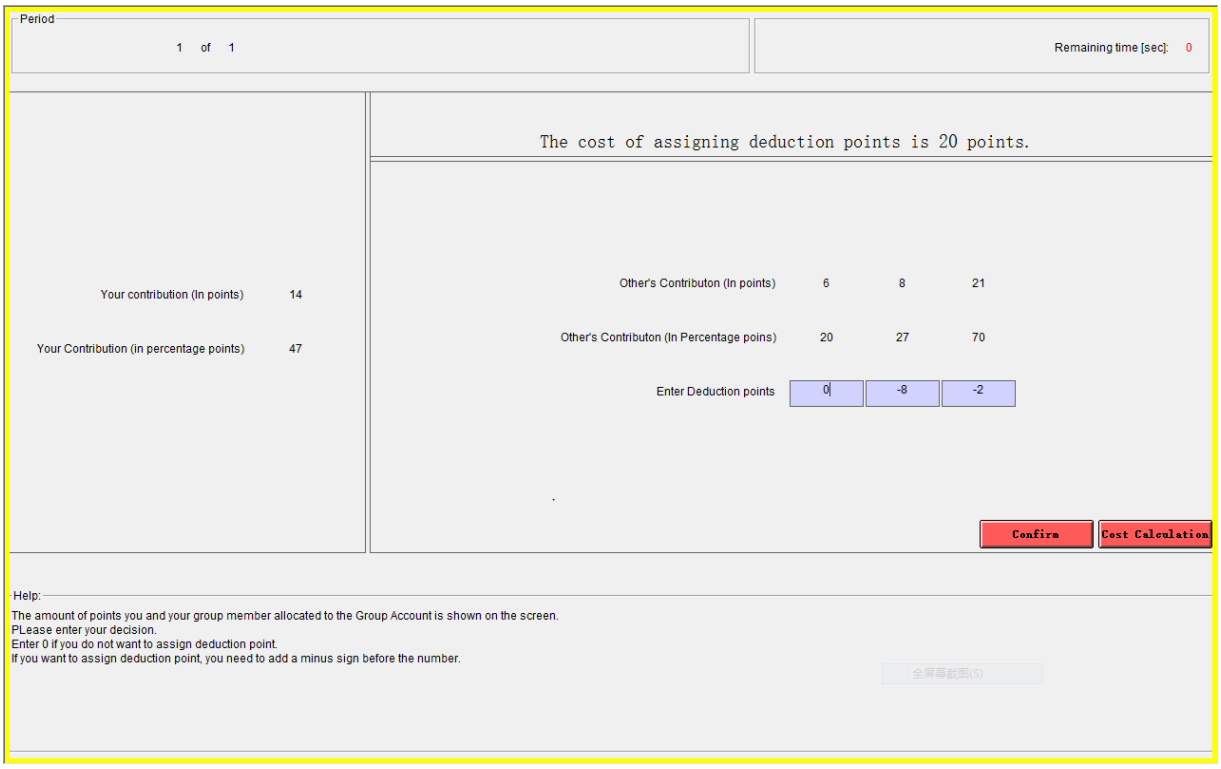
20 - (Your allocation to the Group account) + 0.4 x (Total points allocated to Group Account by all group members)

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:

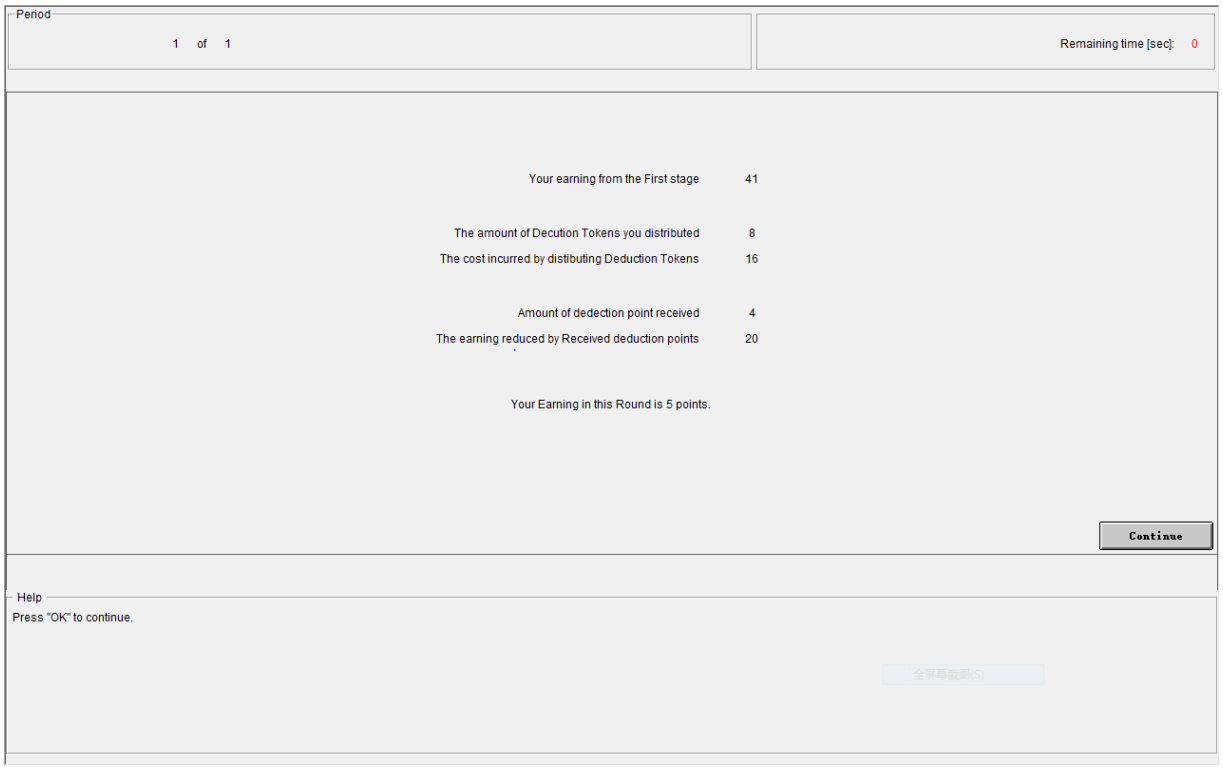
Your earnings at the end of the second stage = income per period

= **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?.....