Cooperation, Institutions and Sustainable Development: empirical evidence from China

Zihan Nie

Thesis committee

Promotors
Prof. Dr E.H. Bulte
Professor of Development Economics
Wageningen University & Research

Prof. Dr Q. Tu
Professor at School of Economics and Resource Management
Beijing Normal University, China

Co-promotor
Dr N.B.M. Heerink
Associate professor, Development Economics Group
Wageningen University & Research

Other members
Prof. Dr S. Feng, Nanjing Agricultural University, China
Dr M.J.J. Handgraaf, Wageningen University & Research
Prof. Dr E.C. van Ierland, Wageningen University & Research
Prof. Dr P.J.M. Oosterveer, Wageningen University & Research

This research was conducted under the auspices of the Wageningen School of Social Sciences.

Cooperation, Institutions and Sustainable Development: empirical evidence from China

Zihan Nie

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Tuesday 3 July 2018
at 11 a.m. in the Aula.

Zihan Nie
Cooperation, Institutions and Sustainable Development: empirical evidence from China, 162 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2018)
With references, with summary in English

ISBN 978-94-6343-784-4
DOI https://doi.org/10.18174/451360

To my family

CONTENTS

Chapter 1 Introduction

1.1 Overview

Cooperation is vital to human society. A simple definition of cooperation is that one individual pays a cost for the benefit of another (Rand and Nowak 2013). Historically, cooperation is one of the underlying reasons that explain the evolutionary success of human species (Nowak and Highfield 2011). The scale to which humans cooperate is considered as the most distinctive feature of humans or *Homo sapiens* as a species (Boyd and Richerson 2005). For the sustainable long-term development of human society, cooperation is also essential in solving social dilemma situations ranging from local issues such as the management of local common pool resources to global issues such as climate changes. Understanding what affects cooperation and how to improve cooperation is crucial in this respect.

This thesis examines factors that affect cooperation among individuals in the context of China. It starts with an assessment of the food certification system for agricultural products, where cooperation among a larger number of farmers is crucial for its success. Upon finding that free-riding incentives are behind the ineffectiveness of the eco-certification system in reducing agro-chemical use, this thesis continues to explore how various contextual and institutional factors affect cooperation. First, it looks at the role of contextual factors, specifically, the role of resource scarcity in shaping cooperation in the context of irrigation agriculture. Then, it moves on to the role of punishment and reward institutions in improving cooperation with particular attention to people's institutional preferences. And last, this thesis examines the effect of leading-by-example and leadership legitimacy on promoting cooperation.

While each of the Chapter 2-5 in this thesis is a stand-alone research article answering a specific research question, they all revolve around the topic of voluntary cooperation. The existence of free-riding incentives and the lack of cooperation hamper the effort of achieving sustainable development. We set out to explore several ways to overcome free-riding incentives and achieve better cooperation among individuals.

1.2 Cooperation for sustainable development

Increasing and imminent resource and environmental problems pose serious challenges for sustainable development of human society. Yet the difficulty of solving these problems lies in the "social dilemma" feature of such problems, where social interests are at odds with individual interests. While society as a whole and everyone in it can benefit if they cooperate by adopting the desirable behavior, each individual has incentives to defect and free-ride on the effort made by others for his or her own benefits. This situation is especially true for large-scale collective action, where a large number of stakeholders are involved and it is very difficult to monitor individual behavior. Thus, cooperation is needed to solve such social dilemmas and failing to cooperate will jeopardize the efforts to achieve sustainable long-term development.

One important case that will be examined in this thesis is the food certification in China. The certification of food, such as Organic Food and Green Food, aims to both increase farms' income and achieve green agricultural development by reducing agrochemical consumption in agricultural production. Intensive application of chemical fertilizers and pesticides has caused contamination of soil, surface water, groundwater and farm products, and it is the major source of rural non-point source pollution in China (Le et al. 2010, Qu et al. 2011, Sun et al. 2012, Guo et al. 2010). Under certification, farmers pledge to adopt more environmental friendly ways of agricultural production. The success of the certification system is built on the premise than farmers reduce their agrochemical use in food production. However, given the price premium obtained from the certification and the difficulty to monitor individual farmers' agrochemicals use on a daily basis, there exist monetary incentives for an individual farmer to use more agrochemicals for higher yields. Voluntary cooperation from farmers in the form of complying with the required practices is thus crucial for the certified food production. However, despite the fact that certified production has developed into a significant segment of China's agricultural sector[*], quantitative studies on farmers' behavior under certification in China are rare. To our best knowledge, Chapter 2 is the first attempt to evaluate the effect of certified food production on agrochemicals use in China on a large scale. Using panel data from six provinces over the period of 2005-2013, we show that on average certified food production does not reduce agro-chemical use as it is supposed to, and that the monetary incentive for free-riding plays a role in explaining the ineffectiveness. The existence of free-riding behavior, and

[*] In 2014, the environmental monitoring area of crop land under Green Food certification was 207.85 million mu (CGFDC, 2015), equal to 10.25% of the total arable land area in China.

thus the lack of cooperation, among Chinese farmers is likely to jeopardize the sustainability and long-term success of the certification system and thereby the general environment.

1.3 To improve cooperation

Lack of cooperation can significantly undermine effort towards sustainable development. Thus, it is important to understand what factors affect voluntary cooperation decisions and how to improve cooperation despite the existence of free-riding incentives.

A recent but growing strand of literature emphasizes the role of contextual factors in shaping preferences, culture and social norms. For example, among other things, market institutions (Bowles 1998), political regimes (Brosig-Koch et al. 2011, Kuhn 2013, Heineck and Sussmuth 2013), production technologies (Leibbrandt, Gneezy, and List 2013, Alesina, Giuliano, and Nunn 2013, Talhelm et al. 2014), workplace organization (Carpenter and Seki 2011, Gneezy, Leibbrandt, and List 2016) and even prenatal exposure to stressful environment (Cecchi and Duchoslav 2018, Duchoslav 2017) can shape social preferences and cooperative norms.

Growing resource scarcity is a critical contextual factor that provides incentives and constraints that guide people's behavior. Therefore, it can shape preferences and social norms as well (e.g. Sarangi, Jha, and Hazarika 2015, Prediger, Vollan, and Herrmann 2014). While many economic studies manipulate resource scarcity using experimental methods (Pfaff et al. 2015, Blanco, Lopez, and Villamayor-Tomas 2015, Gatiso, Vollan, and Nuppenau 2015) or focus on the impact of naturally occurring shocks on resource availability (Cassar, Healy, and Von Kessler 2017), few have examined the effect of relatively long-term exposure to resource scarcity on cooperation. In Chapter 3, we examine the impact of long-term exposure to resource scarcity on cooperation in a canal irrigation system in northwest China and show how resource scarcity helps to foster better cooperation among Chinese farmers.

The institutional environment can also influence cooperation. Institutions are humanly devised to constraints to construct political, economic and social interaction (North 1991). Appropriate institutions can help to reduce free-riding behavior and thus improve and maintain cooperation (e.g. Ostrom 1990). While formal institution is often exogenously imposed on local people especially in non-democratic countries, people exhibit preferences for certain institutions over alternative institutions when allowed to choose (Botelho et al. 2005, Ertan, Page, and Putterman 2009, Gurerk, Irlenbusch, and Rockenbach 2006, Dickinson, Masclet, and

Villeval 2015, Gurerk 2013, Sutter, Haigner, and Kocher 2010). Thus, it is important to know what the implications on cooperation behavior are when the implemented institutions do not match people's institutional preferences and what explains peoples' particular institutional preferences. Yet, few studies have paid attention to this topic. In Chapter 4, we examine how people's institutional preferences interact with assigned institutions in a social dilemma situation and explore the link between people's institutional preferences and their social preferences profiles.

One particular institutional arrangement that helps solve social dilemmas is the sequence of decision-making. In a sequential decision-making scenario, the first mover or the leader can influence others with his or her behavior. By acting as role models, leaders can shape expectations of their constituency and induce them to engage in socially desirable activities. This phenomenon is often referred to as leading-by-example and it has been proven to be an effective instrument in promoting desirable cooperative behavior and overcoming social dilemma situations (e.g. Gachter et al. 2010, Guth et al. 2007, Levati, Sutter, and van der Heijden 2007, Potters, Sefton, and Vesterlund 2007, List and Lucking-Reiley 2002). Upon having a leader, the legitimacy of the leader also has an impact on both leaders' and followers' behavior (Levy et al. 2011, Brandts, Cooper, and Weber 2015, Reuben and Timko 2017). Although many studies have examined the effect of leading-by-example and leadership legitimacy on cooperation, these studies do not separate the effect of sequential move from the effect of legitimacy on cooperative behavior, especially on leaders' behavior, in a satisfactory manner. Therefore, in Chapter 5, we study the impact of leading-by-example and leadership legitimacy on voluntary cooperation with a special experimental design with a special attention to leaders' behavior.

1.4 Objectives

Economic and social interactions heavily depend on cooperation. Cooperation is also of fundamental importance for the sustainable development of human society when facing the challenges such as sustainable resource management and climate changes. Understanding what affects voluntary cooperation and how to improve cooperation is therefore of crucial importance.

The overarching objective of this thesis is to improve our understanding of what factors affect cooperation and to identify the underlying mechanisms of voluntary cooperation. Specifically, the chapters in this thesis answers the following research questions:

1. Does food certification reduce agrochemical use in China?

2. How does long term exposure to resource scarcity affect cooperation?

3. How do punishment and reward institutions and people's institutional preferences affect cooperation?

4. How do leading-by-example and leadership legitimacy affect cooperation?

Chapter 2 is where we identify that lack of cooperation is one of the reasons why the food certification system fails to achieve its main goal. It serves as a primer for Chapter 3, 4 and 5. In Chapter 3, 4 and 5, we examine different mechanisms that affect voluntary cooperation in various settings.

1.5 Methodology

Traditionally, economic behavior is measured using surveys or other observational methods in the field. The data is then analyzed through econometric and statistical methods to identify patterns and relationships, especially causal relationships. Yet, behavior reported or observed in the field are often confounded by many factors and its empirical analyses may suffer from biases and endogeneity issues. Variations in the key factors that are supposed to affect behavior also may seriously suffer from endogeneity problems that undermine causal claims of the findings. Thus, experiments have become increasingly popular in economics literature since 1960s for the purpose of better measurement and better identification (Falk and Heckman 2009, Holt 2007). Harrison and List (2004) has proposed an insightful typology for various experiments used in economics studies. This thesis involves three particular types of experiments: lab experiments, lab-in-the-field experiments and natural experiments.

Compared with traditional methods, lab experiments provide controlled variations in controlled decision-making environments. The controlled environment grants lab experiments advantages in measuring preferences and identifying causal relationships. Behavior observed in a lab experiment can be used as an indicator of preferences, as lab experiments can provide an incentivized, anonymous/private and abstract decision-making environment, while behavior observed in the field and self-reported preferences are often confounded by external factors

and may suffer from self-reporting biases due to e.g. exaggeration and social desirability. Moreover, in lab experiments, experimenters can manipulate the variations in one particular aspect while holding other conditions constant, which allows them isolate the causal effect of interest. Thus, lab experiments are very useful in testing economic theories, especially in testing the effects of factors that cannot be easily manipulated in reality or have no real-life counterparts.

While lab experiments provide the most controllable environments for identifying the causal relationships and providing profound insights for economic theories, there exists skepticisms about the possible lack of "realism" or external validity due to the extensive use of college students as experimental subjects and the lack of real-life contexts (Levitt and List 2007). While most criticisms do not seriously jeopardize the efficacy of lab experiments (Falk and Heckman 2009, Falk, Meier, and Zehnder 2013), the lack of context may influence the results if we have reasons to believe that a specific context in the field matters for people's decision-making processes and the behavior under that specific context is the one of particular interest. If the direction or the size of the causal relationship of interest varies in different contexts, results from a typical lab experiment with college students are not enough. In such cases, putting the experiments in the proper context is important.

Lab-in-the-field or (artefactual field) experiments aim to transplant the conventional lab experiments to the field. They allow researchers to recruit "real" people from the population and the context of interest, but still provide tightly controlled decision-making environments to isolate the effect of interest. Researchers thus are able to study the causal effect of interest among the population of interest.

Lab and lab-in-the-field experiments are powerful instruments in testing causal relationships in various settings, but they also have their limits. Not all variations of interest can be easily or ethically manipulated by the researchers in lab settings. Natural experiments, on the other hand, take advantage of naturally occurring events that are independent from the outcome of interest and treat the differences in the exposure to such events as the exogenous variations or "treatments" to study the causal relationship of interest. As natural experiments happen in a natural context, they reveal peoples' behavior in natural settings with natural constraints. However, the key feature of natural experiments is also the major disadvantage of natural experiments. Naturally occurring events means researchers cannot choose what the "treatments" are and when, where and how they are implemented.

This thesis combines data from experiments and conventional household surveys to examine the role of cooperation, the determinants of cooperation and how to promote cooperation. Specifically, Chapter 2 uses conventional rural household survey data and conventional econometrics methods to evaluate the effectiveness of certification in reducing agro-chemical use. In Chapter 3, 4 and 5, we study cooperation with experiments within the framework of public good games (Ledyard 1995). Chapter 3 uses a lab-in-the-field experiment to measure the willingness of cooperation and uses a quasi-natural experiment with exogenous variation in resource scarcity to study the impact of long-term exposure to resource scarcity on cooperation. Chapter 4 uses a lab-in-the-field experiment to study the interaction between institutional preferences and imposed institutions and the role of social preferences. Chapter 5 uses a lab experiment with college students to study the effect of leading-by-example on cooperation and the role of leadership legitimacy.

1.6 Outline

The rest of this thesis is organized as follows. Chapter 2 examines the effect of obtaining certification on agrochemical use in China using rural household survey data. Chapter 3 studies the effect of long-term exposure to resource scarcity on cooperation in the context of irrigation agriculture in northwestern China using a quasi-natural experiment approach. Chapter 4 examines the role of institutions and institutional preferences in promoting voluntary cooperation. Chapter 5 examines the effect of leading-by-example and leadership legitimacy on promoting voluntary cooperation. Chapter 6 concludes with discussion on policy implications, limitations and recommendations for further research.

Chapter 2 Certification and agrochemical use

evidence from the food certification system in China

Abstract: Food certification, such as green food or organic food, is supposed be able to reduce the overuse of agro-chemicals. Yet, little research has examined the impact of certification on fertilizer and pesticide use. Based on a panel data set from 2005 to 2013 and an additional household survey designed by the authors in 2014, this study examines the effect of adopting certified food production on chemical fertilizer and pesticide use in agriculture in six different provinces. Overall, we do not find significant effects of certified food production on either chemical fertilizer or pesticide consumption among Chinese farmers. The effects are heterogeneous across villages, but the heterogeneous effects show no clear pattern that is consistent with different types of certification. Our findings are robust to the use of alternative panel structure and certification indicators. The lack of knowledge about certification, the price premium and differences in regulation enforcement across regions may explain why we do not find negative effects on agrochemical use.

This chapter is based on:

Zihan Nie, Nico Heerink, Qin Tu & Shuqing Jin, Does Certified Food Production Reduce Agro-chemicals Use in China?, *China Agricultural Economic Review*, forthcoming.

2.1 Introduction

Agricultural production in China experienced rapid growth during the past few decades. The widespread adoption of chemical fertilizer and pesticide has made an important contribution to this growth. Yu and Zhao (2009) provides a literature review on the source of agricultural growth and identifies that "fertilizer is the largest contributor in physical inputs to agricultural growth in China". China now is the largest chemical fertilizer and pesticide consumer in the world.[2] Chemical fertilizer consumption has increased from 25.90 million tons in 1990 to 59.12 million tons in 2013, while pesticide consumption has risen from and 0.73 to 1.80 million tons during the same period (China Rural Statistical Yearbook 2014). According to World Bank data, the average chemical fertilizer consumption in China increased from 325.8 kg to 364.4 kg per hectare of arable land between 2005 and 2013. In France and the United States, and in densely populated Japan after 2008, fertilizer consumption per hectare was considerably lower than in China. Moreover, fertilizer consumption either declined or remained stable in those countries while it increased steadily in China. Pesticide use per hectare is also high in China. According to FAO statistics, pesticide consumption in China is much higher than pesticide consumption in major crop exporters such as the US and France and has surpassed pesticide consumption in Japan since 2007. In contrast to the slightly declining or stable trends in those countries, China has experienced a steady increase in pesticide consumption per hectare in recent years.

The relatively high, and increasing, multiple cropping index in China can at least partly justify the large agrochemical consumption.[3] But the high input intensity *per se* could cause environmental problems and threaten long-term land productivity. In China, chemical fertilizer use efficiency is only around 33% (Cheng, Shi, and Wen 2010, Wu 2011)(Ministry of Agriculture (MOA), 2015a), while pesticide use efficiency is estimated at 35% (MOA, 2015b). Chemical fertilizers and pesticides have been identified as the major source of rural non-point source pollution in China. MOA (2004) claims, "the area of farmland suffering from agricultural chemical contamination has reached 136 million mu[4]". Nitrogen and phosphate washed from agricultural fields are identified as the major source of eutrophication of surface waters and shallow groundwater in China (Le et al. 2010, Qu et al. 2011, Sun et al. 2012). The use of large quantities of agrochemicals also contributes to the acidification of farmlands (Guo

[2] Data source: FAO data base.
[3] The multiple cropping index is defined as the total annual sown area of crops / the total cultivated area *100%. It increased from 120.1 in 1998 to 134.3 in 2012 (Xie and Liu 2015).
[4] 15 mu = 1 ha.

et al. 2010). Intensive application of pesticide has caused contamination of soil, surface water, groundwater and farm products (Sun et al. 2012).

Researchers and policy makers in China have become aware of the downside of the increasing agrochemical consumption and are seeking ways to promote sustainable agricultural development. For instance, in February 2015, the Ministry of Agriculture launched the "Chemical Fertilizer and Pesticide Consumption Zero-Growth Operation", aimed at reducing the growth rate of chemical fertilizer and pesticide consumption to zero, and improving the use efficiency to at least 40%, by the year of 2020.

The general public has also become increasingly aware of the danger of heavy agrochemical use as well, which has resulted in a rising demand for environmental friendly and safe food (Ortega et al. 2015, Thogersen et al. 2015, Yin et al. 2017, Yin et al. 2010, Yu, Gao, and Zeng 2014). However, consumers cannot observe the agricultural practices that produce the food they consume. One way to bridge this information gap is certification. Green Food, Organic Food, Non-hazard Food and Geographic Indication Product are the typical types of certified products in China. Food certificates are issued and controlled by government agencies or qualified third parties. Production of certified food in China aims to meet consumers' demand for high quality food, increase farmers' incomes through the price premium they receive for high-quality food and to reduce agricultural pollution through lower agrochemical consumption. Restrictions on chemical inputs use are usually the crucial farm-level requirement for food certification.

Starting in the 1990s, certified food production has developed into an increasingly important segment of China's agricultural sector. For example, in 2014, the environmental monitoring area of crop land under Green Food certification was 207.85 million mu (Centre 2015) (CGFDC, 2015), equal to 10.25% of the total arable land area in China.[5]

Despite the growing popularity of certified food production in China, the available international scientific literature on its development and impact is limited and mainly qualitative, reviewing the development of certified food production, pointing out existing problems and providing policy advices (e.g. Lin, Zhou, and Ma 2010, Oelofse et al. 2010, Sanders 2006, Scott et al. 2014, Sheng et al. 2009). The available Chinese scientific literature is relatively rich, but largely focuses on marketing strategies, value chain management and

[5] The arable land area equals 135.1634 million hectares (2.027 billion mu), according to recent statistics released by the Ministry of Land and Resources (2015).

development strategies, using evidence from either macro-level data or specific case studies (e.g. Han 2010, Wang, Zhao, and Du 2009). Changes in actual production practices and the environmental outcomes of certified food production have not yet received adequate attention. Much of the certified food production in China is carried out by smallholders. Since it is hard to observe and control the behavior of large numbers of small farmers, it remains unclear to what extent certified food production has actually reduced farmers' chemical fertilizer and pesticide consumption and what environmental benefits it has brought so far.

The available literature on other countries is limited as well. Among the studies that did examine production practices of organic farming, Kleemann and Abdulai (2013) find that organic farming increases agro-ecological practices among pineapple producers in Ghana; Blackman and Naranjo (2012) find that organic farmers use fewer agrochemicals than conventional farmers in Costa Rica; Beuchelt and Zeller (2011) report zero chemicals use among organic farmers in Nicaragua. On the other hand, Gambelli et al. (2014) find evidence of non-compliance occurring in Italy and Germany. Given the weak enforcement ability in low- and middle-income countries, it is likely that non-compliance also exists and may be even more prevalent in such countries. We certainly should not take lower chemical input use among certified food producers for granted. This also holds for China, where the restrictions posed on agrochemical use differ between different categories of food certificates (as will be discussed in Section 2.2).

A few Chinese studies provide some evidence that farmers do not fully conform to the certification requirements. For example, Zhang (2012) examines the quality control behavior of green vegetable farmers in Sichuan province and finds that farmers' pesticide use often differs from the level required by the contracting firms. Zhou and Xu (2008) find that pesticide overuse and use of forbidden pesticides is quite common among farmers engaged in Non-hazard Food production in the city of Nanjing. And Li, Zhu, and Ma (2007) find that Non-hazard Food certification has no impact on farmers' levels of pesticide use in Nanjing. The evidence of non-compliance in these studies is obtained for relatively small regions. It points to the urgent need for a more systematic examination of the issue.

In this study, we use panel data from six provinces to examine the impact of certified food production on chemical fertilizer and pesticide consumption in China. Our sample covers different agro-ecological regions across the country and therefore provides a relatively complete picture of certified food production in China. On average, we do not find evidence of lower agrochemical use in certified food production. The effects are heterogeneous across

villages, but the heterogeneous effects show no clear pattern that is consistent with different types of certification. Our results are robust to the use of alternative panel structure and certification indicators. Our findings suggest that the current product certification system is not very successful in reaching its environmental goals. Stricter inspections and enforcement may be needed in the near future to make a contribution to the reduction of non-point source pollution.

The rest of this chapter is organized as follows. Section 2.2 reviews the development of different types of certified food products and presents our hypotheses. Section 2.3 introduces our empirical strategy and Section 2.4 describes the data set; Regression results and their interpretations are presented in Section 2.5; Section 2.6 provides some discussions about our findings; and Section 2.7 concludes.

2.2 Food certification in China and hypotheses

In China, different types of certified products have different requirements in terms of chemical fertilizer and pesticide use. Non-hazard Food has the lowest requirements, while Organic Food has the highest. Green Food lies in between. Geographic Indication Product certification is not introduced to reduce agrochemical consumption, but it complies with at least the Non-Hazard Food standards.

In 2001, MOA launched the "Non-hazard Food Project" and the formal certification was introduced in 2003. Non-hazard Food certification is created as a way to provide basic food safety. Use of highly toxic pesticide is forbidden and pesticide residuals, heavy metals and microorganism contents need to be below national standards.

Green Food certification started as early as 1990. The China Green Food Development Centre was founded in 1992 to oversee the certification process. There used to be two different types of Green Food certificates: Green Food AA and Green Food A. Green Food AA has very similar standards to organic food, requiring that no chemical fertilizers, pesticides and other chemical inputs are used. It is not widely recognized as organic food and was abolished in 2008, shortly after the official introduction of Organic Food certification. Green Food A is the normally recognized green food, allowing limited use of chemical fertilizer, pesticide and other chemicals; In this study, the term "green food" refers to Green Food A. The Green Food certificate is also valid for three years, and there are random inspections every year.

The export of organic food from China started in the early 1990s, with certification by foreign agencies. The national organic food standards and certification were introduced in 2005. Consistent with the common practice around the world, no agrochemical inputs are allowed in the production process. The certificates are valid only for one year and thus need to be renewed annually.

Geographic Indication Product certification was introduced in 2005 to protect high quality products from certain specific regions. The certification is managed by the Agricultural Products Quality and Safety Centre. Although this certificate does not directly aim to reduce the use of agrochemicals, the inspections and tests follow the same national standards as Non-hazard Food. The Geographic Indication Product certificate is also valid for three years with annual random inspections.

From the technical requirements on limiting agricultural chemical inputs in order to obtain ecological certificates, we can derive a number of hypotheses. Our main hypothesis is as follows: On average, certification of food production has a negative impact on chemical fertilizers and pesticides consumption. Based on this main hypothesis, we further form two sub-hypotheses. Our first sub-hypothesis is that the impacts of certifications are stable over time. This requires that the annual inspections are effective and that producers do comply with the chemical input restrictions. And our second sub-hypothesis that will be tested is that Organic Food certification has the strongest negative impact on agrochemical use, followed by Green Food certificates, and the impact of Non-hazard Food and Geographic Indication Product certification is the smallest. In the next section, we explain how these hypotheses are tested in this study.

2.3 Empirical strategies

To identify the impact of food certification on farmers' agrochemical use, we utilize the panel structure of the data set (see next section) and estimate fixed effects models. A fixed effect model helps to control for potential endogeneity problems related to unobserved time-invariant household-specific effects (u_i).

The main reduced form models we use in this study can be formulated as:

$$Y_{ijt} = \beta_0 + \alpha_1 Cer_{jt} + \alpha_2 CerDur_{jt} + \beta_2 X_{ijt} + u_i + \theta_t + \varepsilon_{ijt} \quad (1),$$

where Y_{ijt} is the indicator of fertilizer and pesticide consumption of household *i* in village *j* in year *t*; Cer_{jt} represents the certified food production dummy and $CerDur_{jt}$ stands for the duration of certified food production in year *t*; *X* is a vector of covariates; u_i is the unobserved household-specific fixed effects; θ_t is the year-specific effects and ε_{ijt} is the error term.

The dependent variable Y_{ijt} is the chemical fertilizer (pesticide)[6] consumption (in kg/mu), which can be interpreted as chemical fertilizer (pesticide) use intensity. It equals the quantity of fertilizer (pesticide) used in a year divided by the total sown area. Since we do not have crop-level agrochemical consumption information for the years before 2010, we use the household as the observation unit. The quantity purchased by a household is used as a measure of total household fertilizer (pesticide) consumption. We use the physical quantity rather than nutrients (active ingredients) content due to incomplete information about the latter.[7] Total sown area was calculated by summing data on the sown areas of 16 different types of crops obtained through the household survey. We use sown area rather than arable land area because it takes multiple cropping practices into account.

Village-level certified food production indicators are used as key explanatory variables. The certified food production dummy variable, Cer_{jt}, indicates whether village *j* had certified food production in year *t*. The dummy equals 1 for all villagers in a village if this village has certified food production in year *t*, and 0 otherwise. The duration variable, $CerDur_{jt}$, indicates the duration (in years) since certification started. We first include only the dummy indicator, Cer_{jt}, to test the main hypothesis that certified food production reduces agrochemical use. Then we further test the sub-hypothesis that the impact of certified production on agrochemical use is stable over time by adding the duration indicator $CerDur_{jt}$. We expect α_1 to be significantly negative and α_2 to be equal to zero.

We choose to use village-level indicators instead of household-level certification indicators for several reasons. First, besides the direct effect of certification on the participants, there might be spillover effects on the non-participants as well. Village-level indicators will

[6] Pesticides include insecticides, herbicides, fungicides, bactericides, and other chemicals meant to reduce potential yield losses.

[7] Information on the nutrient content of fertilizers and percentage active ingredient of pesticides is not available in the RFOP survey data sets. We use the price of fertilizer (pesticide) per kg as an explanatory variable in the regressions to control for differences in quality of fertilizer and pesticide, assuming that the price is an appropriate indicator of the nutrients (effective ingredients) content level. The price per kg is calculated by dividing the total cost of fertilizer (pesticide) by the total physical quantity. We also ran regressions with the monetary value of fertilizer (pesticide) as dependent variable. The results, which can be obtained from the first author upon request, are very similar and do not lead to different conclusions.

take the potential spill-over effects into account. Second, participation in certified food production is a self-selection process. Farmers' decisions of joining or quitting and the timing of these decisions are all endogenous. Given the data we have, these decisions are untraceable and we are not able to deal with these endogeneity issues. One concern is that using village-level indicators would cause biased estimates as not all farmers in certified villages participate in certified food production. However, if certification reduces agrochemical consumption of participants and there is no spill-over to non-participants, we should still expect to find a statistically significant negative effect of the village-level certification variable despite potentially underestimating the size. If there is a spill-over effect that also reduces agrochemical use of non-participants, then our estimates capture the overall effect at the village level. Therefore, if certification affects agrochemical use in the hypothesized direction, the potential bias only affects the magnitude, but not the sign of our estimates. The participation rates in certified villages are relatively high. The average participation rate in certified villages in our sample is about 57%. Thus, the size of the potential bias is limited.

We also use time leads of Cer_{jt} as alternative key explanatory variables. The underlying logic is that, to obtain certification, farmers may have to make preparations in advance. If they already reduce their chemical consumption before getting certified, results from equation (1) may underestimate the negative impacts. We use 1-year and 2-year leading variables respectively. Replacing Cer_{ijt} in equation (1) with Cer_{jt+1} or Cer_{jt+2}, we have the following model:

$$Y_{ijt} = \alpha_0 + \alpha_1 Cer_{jt+1(2)} + \alpha_2 X_{ijt} + u_i + \theta_t + \varepsilon_{ijt}. \qquad (2)$$

Furthermore, to test heterogeneous effects of different types of certification, we construct village-specific certified food production indicators and test whether the coefficients of villages with different certificates follow the pattern described in the second sub-hypothesis. Instead of having one overall indicator, there are seven variables for each of the seven villages:

$$Y_{ijt} = \beta_0 + \beta_{11} Cer_{j1t} + \beta_{12} Cer_{j2t} \ldots\ldots + \beta_{17} Cer_{j7t} + \beta_2 X_{ijt} + u_i + \theta_t + \epsilon_{ijt}. \qquad (3)$$

In all specifications, the vector of control variables *X* consists of other factors besides certification that may affect farmers' fertilizer/pesticide input decisions. It includes household head characteristics, agricultural assets and endowments, other household characteristics, agrochemical prices, crop structure, village characteristics and year dummies. Specifically, household head characteristics include self-evaluated health and agricultural training dummy; agricultural assets and endowments include sown area and its squared term, household labour

force per mu[8], draft animals per mu, large agro-tools per mu, machinery power per mu, proportion of land covered with greenhouses in arable land and proportion of irrigated land in arable land; other household characteristics include computer and television ownership dummies, agriculture as the main income source (dummy variable) and amount of non-cropping income; agrochemical prices include fertilizer price and pesticide price (yuan/kg); the fractions of each type of crops in the sown area are used as indicators of crop structure[9]; village characteristics include two dummy variables: whether the village administration spent money on providing extension services and whether the village administration provides any monetary support for grain production.

Although we hypothesize in Section 2 that certification reduces agrochemical use, we test the null hypothesis of no effect in the analysis. In principle, the effect of certification could go in either direction. Producers could increase agrochemical use if proper monitoring by regulation bodies is missing and if the prices of certified products are higher. We therefore test the null hypotheses of zero effect in the econometric analysis.[10]

2.4 Data

The basic data set used in this study is the Rural Fixed Observation Points (RFOP) survey data set collected by the Research Center for Rural Economy (RCRE), Ministry of Agriculture. The RFOP survey is a national-wide longitudinal survey of rural households and villages. The first survey started in 1986; it is conducted annually except for a few interruptions. The latest available wave in 2013 covers 23,000 rural households living in 360 villages in 30 provincial level administrative units. The survey provides information on rural household characteristics, land use, labor use, agricultural production, income, expenditure, *etc*. Its panel nature allows us to track the changes in fertilizer and pesticide use over time. For a more detailed discussion on the sampling method, see Benjamin, Brandt, and Giles (2005) and Yao (2011).

[8] Household labor force is defined as household members who are 16 or older, not disabled and not students.
[9] The RCRE survey divides crops into 3 main categories: staple crops, cash crops and horticultural crops. Staple crops include 6 subcategories: wheat, rice, corn, soy bean, tuber, other staple crops. Cash crops include 8 subcategories: cotton, oil crops, sugar crops, flax, tobacco, mulberry, vegetable and other cash crops. And horticultural crops comprise fruit and other horticultural crops.
[10] Given the potential misclassification of using village-level certification indicators, we cannot rule out the possibility that two-side tests may fail to statistically detect a positive effect of certification on agrochemical use. Yet, this issue does not affect our main conclusion that we do not find evidence of lower agrochemical use in certified food production.

The regular RCRE survey does not provide any information on certified production. To collect relevant information, the RCRE added a questionnaire specifically focusing on certified production to the regular survey in a 6-province subsample in the summer of 2014.[11] The additional survey provides information on which village has adopted certified food production at the time of survey and when it started. The questions on certified food production refer to Green Food, Organic Food and Geographic Indication Product. Since Geographic Indication certification only started in 2005 and none of our sampled villages adopted certified production before 2007, we restrict the research period to 2005-2013. We build our certified food production indicators in a retrospective way, based on whether they had certification in 2014 and when it started.

We confine our analysis to the villages that were included in the additional 6-provinces survey held in 2014 and combine the information from that survey with information collected for the same villages in earlier RFOP surveys held by the RCRE. After dropping observations with incomplete information and households that were interviewed for the RFOP survey in less than three years, we have 34,569 observations from 4,830 different households in 72 different villages in our working sample. Out of these 72 villages, 7 villages had certified food production at the time of the additional survey. Among these 7 villages, two started certified production in 2007, one in 2008, one in 2009, two in 2010 and one in 2011. The certified products they were producing include staple food (rice), vegetables (Chinese yam), tea and fruits (apples and kiwi fruits). The number of households in our sample in each year ranges from 3,477 to 3,996 (see Table 1). Only 1,757 households were interviewed in all nine years, but there are many households that are missing for only one or two years. To make maximum use of the available information, we use an unbalanced rather than a balanced panel in our main analysis. We check the robustness of our results in section 5.3 for the balanced panel. [12]

[11] The six provinces are Hebei, Liaoning, Anhui, Fujian, Shaanxi and Yunnan. Due to our budget constraints and existing regulations at the Research Centre for Rural Economy (RCRE), it was not possible to do the survey in all 30 provinces in the RCRE Fixed Observation Point Survey. In consultation with the RCRE, we selected six provinces that represent different agro-ecological and socio-economic conditions prevailing throughout China and where the quality of the RCRE survey data is known to be relatively good.

[12] Information about 969 households (20%) is missing for only one year, and for 638 household (13%) it is missing for only two years. This means that about 70% of the households appear at least seven times in our nine-year time span. In the main analysis, we include households that have information for at least three different years so that we can capture a time trend. The number of observations in each year is rather stable, as can be seen from Table 1. Sample attrition does not monotonically reduce the sample size over time. This means that the panel is unbalanced mainly due to accidently missing information, instead of systematic sample attribution, and that our sample should be able to capture the changes over time well.

Table 1 shows the mean values of the most important variables for the purpose of this research in the nine waves of the survey (see Appendix A for the complete table). The proportion of certified food producers in our sample grew steadily from zero to about 10% in 2013. The chemical fertilizer consumption per mu in our sample shows a fluctuating but slightly increasing trend over the period under study. It increased proximately at rate (12% from 2005 – 2013). Pesticide consumption per mu increased more rapidly over time. The consumption in 2013 is 69% higher than in 2005.

The mean values of household head characteristics and most of the endowment and input variables are relatively stable over time. The sown area fluctuates around 11 mu per household, which is consistent with data presented in China's national statistics.[13] The steady increase of household labor force per mu after 2008 is probably due to the combined effects of more and more younger persons reaching labor force ages and declining mortality rates. Plastic mulching use is more volatile, without a clear time pattern. The share of land occupied by greenhouses in the total arable area is very small in all years. The share of land under irrigation in our sample declined despite an increase from 45.2 percent in 2005 to 52.1 percent in 2013 at the national level (NBS, various years). It is mainly caused by that fact that sample attrition relatively concentrated in Anhui and Yunnan where the shares of irrigation land are higher. The data on crop structure show an expansion of the area grown with maize from 23% in 2005 to 33% in 2013. Partly as a result, the area shares of other major crops decline over time, except for fruits and vegetables. Given the growing importance of migration and other forms of off-farm employment in rural China, it is no surprise that income from non-cropping activities has more than doubled during the period 2005-13. The two price variables are calculated from dividing total cost by the total quantity of fertilizers or pesticides for each household. They therefore partly reflect changes in the composition of fertilizers and pesticides bought by farmers. The price of chemical fertilizers increased about 28% over the entire 2005-13 period. Its annual changes closely resemble fertilizer price changes reported in the China Statistical Yearbooks (NBS, various years). Mean pesticide prices increased more slowly and showed less variation than fertilizer prices for the farm households in our sample.

[13] In 2006, the total sown area equaled 152.1 million ha (NBS, 2008a), while number of agricultural production households stood at 200.16 million at the end of 2006 according to the second Agricultural Census (NBS, 2008b). The average sown area per household therefore equaled 11.4 mu (0.76 ha) at the end of 2006. The relatively large mean sown area in 2006 as compared to the other years reported in the RCRE survey data is probably caused by changes in the sample composition over time.

Table 1. Sample means of selected variables, 2005 - 2013

	2005	2006	2007	2008	2009	2010	2011	2012	2013
Certified producers (%)	0.00	0.00	4.12	5.20	5.78	9.68	11.17	10.22	9.92
Fertilizer use (kg/mu)	67.33	71.19	70.02	66.48	72.34	75.99	74.64	76.88	75.44
Pesticide use (kg/mu)	1.14	1.27	1.28	1.58	1.7	1.53	1.85	1.99	1.93
Sown area (mu)	10.59	13.60	10.60	10.51	10.85	11.09	10.45	10.31	10.99
Household labor per mu	0.62	0.65	0.75	0.78	0.95	0.99	1.11	1.21	1.19
Agro tools per mu	0.19	0.18	0.22	0.22	0.25	0.34	0.23	0.22	0.24
Machinery power per mu	0.23	0.23	0.28	0.25	0.23	0.29	0.44	0.86	0.78
Green houses (% of arable land)	0.59	0.71	0.18	0.22	0.33	0.96	0.20	1.04	0.22
Irrigation (%)	50.87	51.20	55.19	54.61	50.61	49.13	48.90	47.16	46.19
Share of major crops in sown area (%)									
Wheat	16.08	16.55	15.88	14.69	14.06	14.38	14.23	13.72	14.30
Rice	20.05	19.83	20.30	20.07	19.60	18.22	20.28	16.55	17.08
Maize	23.31	23.73	23.27	22.65	26.38	28.97	30.10	33.75	33.08
Soybean	5.85	4.93	4.43	4.71	4.54	3.82	3.07	2.19	2.35
Vegetables	5.28	5.56	5.39	5.87	5.96	5.91	6.69	6.96	6.64
Fruits	8.03	6.66	8.35	9.15	8.59	8.41	8.69	9.14	9.43
Non-cropping income (1,000 yuan)	12.02	13.72	17.19	17.29	18.69	21.31	24.42	25.22	27.98
Fertilizer price (yuan/kg)	1.51	1.5	1.53	1.94	1.78	1.69	1.91	2.03	1.93
Pesticide price (yuan/kg)	23.28	24.67	25.58	26.21	26.92	27.11	27.15	28.34	26.66
No. of villages	70	70	69	70	72	72	69	71	67
No. of observations	3,943	3,909	3,936	3,959	3,996	3,813	3,689	3,847	3,477

Source: Calculated by authors from the RCRE surveys and the 2014 additional survey.
Note: All price and income variables have been deflated to 2005 prices, using national consumer price index data.

2.5 Results

2.5.1 Effects on Household Fertilizer and Pesticide Consumption

We estimate equation (1) with fixed effect regressions. All standard errors are clustered at the village level. The regression results of key explanatory variables are shown in Table 2. Column (1) and (2) present the results for chemical fertilizer consumption. Column (1) includes only the certified production dummy in order to test the average effect of certification, while column (2) includes both the certification dummy and its duration to examine whether the certification effect changes over time. Although the coefficients are large in magnitude (as compared to the mean values in Table 1), they are not statistically significant. Since we expect to find a significantly negative impact, the insignificant results along with the positives signs show no support for our hypothesis that certified food production reduces fertilizer consumption. Nor do the results support the sub-hypothesis that the (negative) impact of certification on fertilizer use remains stable over time.

As for the control variables, it is worth noting that the relation between sown area and fertilizer exhibits a U-shape pattern. Fertilizer consumption per unit area decreases as the sown area increases until it reaches about 450 mu (about 30 hectare). Since the average sown area in our sample is only 11 mu per household, the vast majority lie on the left side of the U-shape curve, suggesting a negative correlation between farm size and chemical fertilizer input. The use of agricultural machineries seems to be complementary to fertilizer use. We find a slightly negative effect of irrigation, probably because given the household fix effects and crop structure, irrigation helps improve fertilizer use efficiency and thus slightly reduce fertilizer input. And farmers in villages that provide staple crop subsidies seem to use less chemical fertilizer. The fertilizer price is negatively related to fertilizer use as expected. Coefficients of the crop structure variables are largely consistent with our expectations that cash crops and horticultural crops usually need heavier fertilizer input than staple crops, although they are not shown in Table 2.[14] Year dummies show a rising trend of fertilizer use, when controlling for other factors included in the model, during the period under study.

The results for pesticide use are reported in columns (3) and (4). The values of R^2 are smaller than those in the fertilizer regressions, suggesting that the explanatory variables have less explanatory power in explaining pesticide use. This is plausible since pesticide use may be more related to certain climate conditions, biological features of crops and outbreaks of diseases or insects, and thus less dependent on other inputs or household characteristics. The average effect of certified food production on pesticide use intensity in column (3) seems to be zero, while column (4) suggests that certified food production has a positive effect on pesticide consumption, although only significant at 10% level. These results imply that having certification either has no impact on pesticide use or even slightly increased pesticide use over time. The hypotheses that certified food production reduces pesticide consumption and that the (negative) impact of certification is stable over time should therefore both be rejected.

[14] Full regression results are available upon request from the first author.

Table 2. Certified food production and agrochemicals use: fixed effects regression results

		Fertilizer		Pesticide
	(1)	(2)	(3)	(3)
Certified production				
Certified production	18.57	4.566	0.413	-0.215
	(1.580)	(0.548)	(0.865)	(-0.537)
Certified duration		5.041		0.226*
		(1.418)		(1.864)
Selected explanatory variables				
Sown area	-0.448***	-0.454***	-0.00849*	-0.00873*
	(-2.688)	(-2.754)	(-1.729)	(-1.804)
Sown area squared	0.000493**	0.000498**	1.01e-05	1.03e-05*
	(2.510)	(2.556)	(1.644)	(1.692)
Household labor per mu	-0.500	-0.520	0.120**	0.119**
	(-0.645)	(-0.670)	(2.427)	(2.403)
Agro tools per mu	-0.624	-0.647	0.0106	0.00961
	(-0.760)	(-0.792)	(0.189)	(0.171)
Machinery power per mu	0.194**	0.188***	0.00568***	0.00541***
	(2.588)	(2.664)	(2.842)	(2.654)
Green house %	0.349***	0.354***	0.0209*	0.0211*
	(3.084)	(3.184)	(1.861)	(1.858)
Irrigation %	-0.0634	-0.0643	-0.000777	-0.000813
	(-1.642)	(-1.662)	(-0.275)	(-0.289)
Agricultural as main income source	4.063*	3.768	0.190	0.176
	(1.676)	(1.641)	(1.278)	(1.188)
Non-cropping income	-0.0585*	-0.0563	-0.00143	-0.00132
	(-1.720)	(-1.627)	(-1.348)	(-1.273)
Fertilizer price	-14.28***	-14.30***		
	(-4.246)	(-4.257)		
Pesticide price			-0.0414***	-0.0415***
			(-7.543)	(-7.493)
Extension service	0.138	0.385	-0.861	-0.850
	(0.0295)	(0.0821)	(-1.606)	(-1.590)
Staple crop subsidy	-9.759*	-9.288*	-0.214	-0.193
	(-1.773)	(-1.771)	(-0.644)	(-0.604)
Crop structure	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes
Constant	87.73***	88.51***	1.328*	1.364*
	(6.176)	(6.214)	(1.773)	(1.807)
Observations	34,569	34,569	34,569	34,569
R-squared	0.101	0.103	0.034	0.034

Note: Only coefficients of selected variables are reported in the table. t-statistics in the parentheses. Standard errors are cluster-adjusted at the village level. ***, ** and * stand for significant level at 1%, 5% and 10% respectively.

In Table 2, we assume that gaining certification at certain year affects agrochemical use in that year and afterwards. However, farmers that plan to obtain a certification may make preparations before applying for it. Therefore, the effect of certification may happen before actually gaining the certification. If this is the case, leading indicators are the more appropriate explanatory variables. Thus, we estimate equation (2) with one- or two-year leading certification dummy as the key variable and report the coefficients of the certification in Table 3. Again, the estimated coefficients are not significantly different from zero. The hypothesized

negative impact of certification on agrochemical use is therefore not supported by the results even when we take the possibility that farmers reduce their agrochemical input before getting the certification into account.

Table 3. Certified food production and agrochemical use: fixed effects regressions with leading indicators

	(1)	(2)	(3)	(4)
	Fertilizer,1-year	Fertilizer, 2-year	Pesticide, 1-year	Pesticide, 2-year
Certified production	6.568	0.0990	0.117	-0.240
	(0.793)	(0.0162)	(0.369)	(-1.044)
Observations	28,048	23,549	28,048	23,549
R-squared	0.114	0.108	0.037	0.033

Note: Control variables are the same as in Table 2. Full regression results are available upon request. t-statistics in the parentheses. Standard errors are cluster-adjusted at village level. ***, ** and * stand for significant level at 1%, 5% and 10% respectively.

2.5.2 Heterogeneous Effects of Certified Food Production

Different types of certificates could have different effects on input use due to the differences in the standards and requirements of each certificate. These heterogeneous effects might be hidden underneath the insignificant impact found in the previous section. Since organic food certification requires using no chemical inputs, green food certification requires limited chemical inputs use and geographic indication certificate exerts no legal constraint on fertilizer use and only forbids the use of highly toxic pesticide, we expect that organic food production has the strongest negative impact, green food weaker and geo-indication food has the smallest impact on agrochemical use (as specified in the second sub-hypothesis).

For this purpose, we construct certified food production indicators for each village and run fixed effects regression as in Table 2 and Table 3 to identify village-specific effects of certified food production for each of the seven villages. Since we can identify which kind of certification each village holds, we can compare the effects of different certifications and test the aforementioned sub-hypothesis. The estimation results are given in Table 4. Only the coefficients for the seven village indicators are shown in the table for conciseness. We mark the type of certification for each village-specific certification indicator in the parenthesis.

The coefficients for the seven village indicators vary a lot. There is only one village in which both fertilizer and pesticide use significantly declined after the certificate was obtained. Four villages experienced increased fertilizer consumption after getting certification, and the magnitudes are quite large in some cases (e.g., in vill1302 and vill6110). Certified food production has no significant effect in two villages and only in one village, we find a negative

effect of certified food production on fertilizer consumption as it is supposed to be. Results for pesticide are similar. There are one village with significantly positive coefficients, five with insignificant coefficients and one with significantly negative coefficients.

For the four villages with Green Food certification, the estimated coefficients for both fertilizer and pesticide use are quite diverse, ranging from large and significant positive values, to insignificant and significant negative values. Therefore, we cannot assess the direction of the impact of green food certification on fertilizer and pesticide input use. The same holds for Geographic Indication certification. Surprisingly, Organic Food certification seems to increase fertilizer use, and have no significant effect on pesticide use, in the two villages that obtained the certification. However, since we only have two villages in our sample and the effects may be case-specific, we don't want to over-interpret these results as evidence against organic food production. To sum up, we find heterogeneous effects across different villages but we do not find evidence supporting the sub-hypothesis that negative effects on agrochemical use increase with the degree of stringency in the type of certificates.

Table 4. Village-specific effects on fertilizer and pesticide consumption

	(1)	(2)
Dependent variable	Fertilizer	Pesticide
Certified production		
Vill1302 (Green)	72.05***	3.615***
	(11.61)	(16.59)
Vill3404 (Green)	0.110	-0.0158
	(0.0474)	(-0.0952)
Vill3415 (Green)	-6.861	-0.388
	(-1.171)	(-0.889)
Vill3511 (Geo & Green)	-26.36***	-0.665***
	(-9.173)	(-4.284)
Vill5302 (Organic)	7.086*	-0.141
	(1.875)	(-0.899)
Vill6110 (Geo)	90.47***	0.645
	(11.03)	(1.519)
Vill6118 (Geo & Organic)	11.13***	-0.339
	(2.790)	(-1.562)
Observations	34,569	34,569
R-squared overall	0.117	0.037

Note: Fixed effect regression results. Control variables are the same as in Table 2. Full regression results are available upon request. t-statistics in the parentheses. Standard errors are cluster-adjusted at village level. ***, ** and * stand for significant level at 1%, 5% and 10% respectively.

2.5.3 Robustness Checks

We use unbalanced panel data in the previous analysis. As a robustness check we also run the same regressions as in Table 2 using a balanced panel for the 1,757 households with

complete information in all the 9 waves of the survey, leaving us 15,813 observations. The regression results, which are shown in panel (A) of Table 5, are consistent with our main conclusions on the effects of certification on fertilizer and pesticide consumption.

Another concern is that large farmers may act differently from small farmers when facing the same production and market conditions. Although the average land holding size is rather small, there are some households that have relatively large land holdings. For example, there are 60 observations with sown area larger than 150 mu (10 hectare) in our sample. Relatively many of these observations are concentrated in the year 2006, which explains the relative peak of average sown area in 2006 shown in Table 1. As another robustness test we restrict our sample to farmers whose total sown area is smaller than 30 mu (= 2 hectares) and redo all the regressions in Table 2. Less than 4% (1,309 out of 34,569) of observations are dropped by this criterion. Yet, our conclusions about the (lack of) impact of certified production on agrochemical consumption do not change when large farmers are excluded from the sample (see panel (B) in Table 5).

In the previous analysis, we use village-level certified production indicators. This means that all households in villages with certified production are regarded as producers of certified crops. When not all agricultural households in a village participated, using village-level indicators may underestimate the negative impact on agrochemicals. Although we do not have household-level information on certified crop production, we may use household-level information on crop production to construct a household level certified production variable. It equals 1 for households who lived in a village that had adopted certified production in a certain year and whose revenues from the certified crop was positive in that year and equals 0 otherwise. The household-level duration variable indicates how many years the household had engaged in certified production. These household-level indicators could help reduce the classification error by excluding seeming non-participants. We rerun the same specifications as in Table 2 and present the key coefficients in panel (C) of Table 5. The duration variable is no longer statistically significant (though larger in value) in the pesticides equation, but the positive impact on fertilizer becomes significant at 10% level. Again, our results do not support the hypothesis that certified food production reduces agrochemical use. Our results are robust to the potential underestimation of negative effects due to using village level indicators.

To address the same concern, we can also use information collected in the additional village survey held in 2014 about how many households participated in certified food production. Combining this information with the total number of households in each village

derived from RFOP data, we calculated the participation rates in the seven villages with certification. We received valid participation rate information for six of these villages. Their participation rates are 23% (vill5302), 37% (vill3511), 57% (6118), 62% (vill1302), 80% (vill6110) and 81% (vill3404), respectively. The average participation rate is 57%. If food certification reduces agrochemical use, we expect to find larger and more statistically significant negative effects in villages with higher participation rates. However, when we combine the participation rates of each village with the village-specific effects presented in Table 4, we do not see such a pattern. Interestingly, the only village showing a significant decrease in agrochemical use has a relatively small participation rate (vill3511: 36.5%), while large increases in the use of agrochemicals can be observed in two villages with a higher share of participants (vill1302: 62.1%; vill6110: 80%). These results suggest that a bias due to differences in participation rates between villages cannot explain our main findings.

So far, we have controlled for crop structures in the regression specifications to assess the impact of certification while taking crop-specific demands for agrochemicals into account. However, as food certification applies to specific crops, crop choices could change as a result of obtaining certification and thereby also indirectly affect agrochemical consumption. Comparing the villages that adopted certified food production (C-villages) with those that never adopted (N-villages) we find that crop structures in C-villages not only were different from crop structures in N-villages in 2005, but also changed in different patterns from 2005 to 2013 (see Appendix B, the shares of certified products such as fruits and rice increased in C-villages while decreased in N-villages). Therefore, in Panel D of Table 5, we exclude crop structure from our regression specifications and compare these results with the results in Table 2 to check whether the indirect effect of certification through changes in crop structures could influence our main findings. Yet, the results here are very similar to our main findings in Table 2. We do not believe that the indirect effect through crop structure changes affects our findings.

Table 5. Robustness checks

	(1) Fertilizer	(2) Fertilizer	(3) Pesticide	(4) Pesticide
(A) Balanced Panel				
Certified production	10.63	10.49	0.367	-0.188
	(1.065)	(1.415)	(0.838)	(-0.516)
Certified duration		0.0495		0.198***
		(0.0320)		(2.745)
Observations	15,813	15,813	15,813	15,813
R-squared	0.130	0.130	0.044	0.044
(B) Smallholding farmers				
Certified production	18.38	4.995	0.388	-0.225
	(1.632)	(0.590)	(0.847)	(-0.557)
Certified duration		4.843		0.222**
		(1.416)		(2.047)
Observations	33,260	33,260	33,260	33,260
R-squared	0.120	0.122	0.037	0.037
(C) Household level				
Certified production	26.90*	3.785	0.839	-0.113
	(1.673)	(0.371)	(1.220)	(-0.230)
Certified duration		8.044		0.331
		(1.494)		(1.545)
Observations	34,569	34,569	34,569	34,569
R-squared	0.104	0.108	0.034	0.035
(D) Without Crop structure				
Certified production	21.44	4.871	0.581	-0.253
	(1.616)	(0.603)	(1.074)	(-0.663)
Certified duration		5.885		0.296*
		(1.511)		(1.873)
Observations	34,569	34,569	34,569	34,569
R-squared	0.081	0.084	0.030	0.031

Note: Fixed effect regression results. Control variables in each panel are the same as in Table 2 (except for Panel D, which exclude crop structure variables). Full regression results are available upon request. t-statistics in the parentheses, based on standard errors cluster-adjusted at village level. ***, ** and * stand for significant level at 1%, 5% and 10% respectively.

2.6 Discussions

We do not find support in our empirical analysis for the hypothesis that producing certified food reduces fertilizer and pesticide consumption. The average impact on fertilizer use is not significantly different from zero and the average impact on pesticide use is even weakly positive. When looking into the heterogeneous impacts of different types of certifications, there are great variations in the direction and magnitude of the effects across villages, with more villages that use more agricultural chemicals in certified production than villages that use less. But there is no clear pattern fitting the different requirements on agrochemical inputs of different certificates to the variation across villages.

One possible explanation for our counter-intuitive findings is that villages that adopted certified food production are villages that already used less agrochemicals than other villages. If farmers started with low agrochemical input, they do not have to change their farming

practices much. To test this, we compared the average fertilizer and pesticide consumption in 2005 between villages that never adopted certified production and villages that adopted certified production at some point of time during our period of observation in Table 6.

Table 6. Comparison of fertilizer and pesticide before certification

	(1)	(2)
	Fertilizer (kg/mu)	Pesticide (kg/mu)
Non-certified villages	66.35	1.12
Certified villages	75.36	1.26
t-statistics	-3.261***	-0.900

Source: Calculated from the RCRE surveys and the 2014 additional survey.
Note: Figures are the average consumption of households in each type of villages. Two-tailed t-tests are used for the mean comparison of the two types of villages. ***, ** and * stand for significant level at 1%, 5% and 10% respectively.

As can be seen from the table, villages that obtained certificates did not have lower levels of fertilizer and pesticide use in 2005. On the contrary, they actually had significantly higher levels of fertilizer use just a few years before they obtained their certificates. Therefore, in the case of China, we have reasons to expect that certification of crop production does not reduce agrochemical inputs use and does not generate environmental benefits.

What may be the reasons why certified production does not reduce chemical fertilizer and pesticide use in China? Farmers' awareness about certified production may be one potential explanation. In the additional household questionnaire in 2014[15], we asked the respondents about their knowledge of each type of certification, using a 1-5 scale: 1 = "never heard of", 2 = "only heard of the name", 3 = "know somewhat", 4 = "know well", and 5 = "know very well". The distribution of the answers is given in Table 7. All respondents were asked to answer these questions no matter whether they were producing certified food or not. Geographic Indication Product is the least well-known certificate. 87 percent of surveyed households answered that they "never heard of" it or "only heard of the name". Green Food is relatively better known, as the corresponding proportion is 72 percent. Furthermore, even among the certified producers, the self-reported knowledge still seems to be limited. Over 50 percent of Green Food producers barely knows anything about it (1 and 2) and the numbers for

[15] Here we use the information collected from the additional household survey in 2014. The data are collected in the same villages as the village survey. However, due to some coding issue, we were not able to match the 2014 additional household survey data with 2005-2013 RFOP survey data well. And the data quality in terms of certified food production is quite poor. Thus, we did not use this household survey in the main analysis and only use it in the discussion section to provide some descriptive evidence.

Geographic Indication Production and Organic Food are even higher. With limited awareness of what they are producing, farmers are unlikely to reduce their agrochemical consumption.

Table 7. Knowledge of different types of certification

	All sample (%)			In certified villages (%)		
	Geo Indic.	Green	Organic	Geo Indic.	Green	Organic
1 = "never heard of"	69.6	36.4	49.0	63.9	38.2	54.1
2 = "only know by name"	17.9	35.8	26.7	11.5	19.7	21.9
3 = "know somewhat"	8.9	20.6	19.7	17.2	37.9	16.4
4 = "know well"	1.6	5.4	3.3	7.1	4.0	7.3
5 = "know very well"	2.0	1.8	1.3	0.3	0.2	0.3
No. of Obs.	4,368	4,364	4,209	366	422	329

Source: Calculated by the authors from the 2014 additional survey.

The price premium coming along with certification may provide farmers with incentives to actually increase agricultural chemical use to obtain higher yields. We have some evidence showing that farmers are not particularly loyal to their cooperators. In the survey, we asked a question about side selling: If there are other firms/individuals that offer 10% higher prices for your certified products, will you sell to them? All Organic Food producers answered "definitely will not" but most of the Green Food (155 out of 214) and Geo-Indication producers (35 out of 66) answered "definitely will". Meanwhile, the average price premium reported in the survey is around 30%, compared with prices of regular crop products. Along with the fact that farming practices of individual farmers are hard to observe and monitor, it is likely that even farmers who know exactly what they sign up for, may not comply.

To formalize this idea, we test how relative importance of certified production affects farmers' agrochemical use. We use the share of the revenues from the certified crop in the total revenues from crop production as a proxy of the relative importance of certified production and add this proxy to the regression specification as an interaction term with certified production dummy. The key results are shown in Table 8. It reveals that certified production reduces agrochemical consumption only for those farmers whose revenues from certified production are just a small share of their total cropping revenues.[16] The more important certified production is for farmers, the more chemical fertilizer and pesticide they use. In our sample, for more than 80% of certified producers, certified production is important enough to incentivize them to use more agrochemicals, rather than less.

[16] The turning point in the share of revenues where the impact of certification on use becomes positive is 41 percent (= 51.15/1.242) for fertilizer and 46 percent (= 2.309/0.0495) for pesticide.

Table 8. Impact on agrochemical consumption and relative importance of certified food production

	(1)	(2)
Dependent variable	Fertilizer	Pesticide
Certified production	-38.37*	-1.657**
	(-1.718)	(-2.228)
Certified production*relative importance	1.116**	0.0427**
	(2.367)	(2.395)
Observations	34,569	34,569
R-squared	0.118	0.037

Note: Fixed effect regression results. Control variables are the same as in Table 2. Full regression results are available upon request. t-statistics in the parentheses. Standard errors are cluster-adjusted at village level. ***, ** and * stand for significant level at 1%, 5% and 10% respectively.

Differences in regulatory enforcement may explain the heterogeneous results across villages. Some villages are doing better in reducing agrochemicals that others. Although we do not have first-hand data on this issue, certification scandals in the news might shed some light on this. One recent example is a report by The Beijing News on the Traceability Code system that allows consumers to look up the origin of the products through the code online, which is particularly relevant to Geo-Indication Products. [17] According to the report, tracing information may be faked and the codes can be purchased at a low price, or even customized at will. Another example is a report on organic certification in 2012.[18] In some places, an organic certificate could be bought for less than a hundred thousand yuan from a certification agency; government inspection and supervision were loose due to unclear jurisdiction or overlapping jurisdiction of different agencies; farmers interviewed admitted not conforming to the production protocols and conforming relied more on self-discipline. Such anecdotal stories may imply that differences in regulation enforcement are a possible explanation for the observed heterogeneous effects of certified production across villages. For villages located in regions with stronger regulation enforcement, certified food production may reduce agrochemical inputs use as it is intended to do. However, for villages in regions with weaker certification enforcement, certified food production may not reduce the use of agrochemicals but merely contribute to larger farm incomes.

2.7 Conclusion

[17] The Beijing News, 2015-6-23. Website link: http://epaper.bjnews.com.cn/html/2015-06/23/content_583285.htm?div=0

[18] China News Service, 2012-8-28. Website link: http://www.chinanews.com/jk/2012/08-28/4137759.shtml

The excessive agricultural chemical consumption in China has drawn more and more attention from the general public and from policy makers in recent years. Certification has been used as an important tool to promote sustainable agricultural development in China. Certified production practices started in the early 1990s, but only began to boom in mid-2000s. Now, it has become an important part of agriculture production in China. However, little is known about the extent to which certified food production brings environmental benefits by reducing the use of chemical fertilizers and pesticides.

Using data from the Rural Fixed Observation Points Survey conducted by Research Center for Rural Economy, and an additional survey specifically focusing on certified food production, we explore the empirical evidence on this question. Using the panel nature of the data, our analysis reveals that certified food production might not be very effective in reducing chemical fertilizer and pesticide consumption. On average, certified food production does not have a significant impact on chemical fertilizer use, while there is weak evidence suggesting that it even increases pesticide use. The effects are found to be heterogeneous across villages. In some villages with certified production the use of agrochemicals does show a decline, but the impacts on both chemical fertilizer and pesticide use in most of villages are significantly positive.

Limited awareness and knowledge about certified food production might be a potential explanation. Little knowledge could easily translate into little change in agricultural practices. Even knowing exactly what they are supposed to do, the price premium could lure farmers to use more agrochemicals to secure or increase yields. Another possible explanation is the weak enforcement of relevant regulations, which may explain the heterogeneous effects across villages that we observed. In regions with careful inspections and strong enforcement, certified food production is more likely to reduce chemical inputs as expected, while in regions with loose inspections and weak enforcement, the price premium of certified food may even stimulate a higher use of agrochemical inputs.

Given the findings of this study, we suggest that further measures should be taken to make sure that the environmental goals of certified food production can be achieved and the reputation of certification in China can be improved. However, the evidence presented in this study should be considered as preliminary evidence that may guide future research in this field. Our study is limited in scale, not able to distinguish between agrochemical use on certified and non-certified crops and only made an explorative analysis of the potential factors explaining the results. The village-level identification strategy that we applied could not be used to directly estimate the effects of certification on agrochemical use by certified food producing households,

nor could it be used to estimate possible spillover effects to other households. Future research is needed to examine the validity of our counterintuitive conclusions through using, for example, data from large-scale surveys specifically designed for this purpose and to rigorously test the potential explanations that we give in our study.

Appendices

Appendix A. Means of Relevant Variables in Analysis

	2005	2006	2007	2008	2009	2010	2011	2012	2013
Fertilizer (kg/mu)	67.33	71.19	70.02	66.48	72.34	75.99	74.64	76.88	75.44
Pesticide (kg/mu)	1.14	1.27	1.28	1.58	1.70	1.53	1.85	1.99	1.93
Certified producer (%)	0.00	0.00	4.14	5.24	5.83	9.66	11.18	10.38	9.61
Agro-training	0.06	0.07	0.08	0.08	0.07	0.07	0.09	0.08	0.08
Health: very good	0.47	0.48	0.49	0.49	0.50	0.50	0.50	0.48	0.48
Health: good	0.37	0.36	0.36	0.36	0.33	0.32	0.34	0.34	0.34
Health: normal	0.11	0.10	0.10	0.10	0.10	0.11	0.09	0.11	0.11
Health: bad	0.04	0.04	0.04	0.04	0.05	0.05	0.05	0.05	0.06
Health: disabled	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Crop sown area	10.59	13.60	10.60	10.51	10.85	11.09	10.45	10.31	10.99
Labor per mu	0.62	0.65	0.75	0.78	0.95	0.99	1.11	1.21	1.19
Draft animals per mu	0.04	0.04	0.04	0.03	0.04	0.03	0.04	0.03	0.03
Tools per mu	0.19	0.18	0.22	0.22	0.25	0.34	0.23	0.22	0.24
Machinery power per mu	0.23	0.23	0.28	0.25	0.23	0.29	0.44	0.86	0.78
Green house (%)	0.59	0.71	0.18	0.22	0.33	0.96	0.20	1.04	0.22
Irrigation (%)	50.87	51.20	55.19	54.61	50.61	49.13	48.90	47.16	46.19
Crop structure (%)									
Wheat	16.08	16.55	15.88	14.69	14.06	14.38	14.23	13.72	14.30
Rice	20.05	19.83	20.30	20.07	19.60	18.22	20.28	16.55	17.08
Maize	23.31	23.73	23.27	22.65	26.38	28.97	30.10	33.75	33.08
Soybean	5.85	4.93	4.43	4.71	4.54	3.82	3.07	2.19	2.35
Potato	3.40	3.47	3.63	3.53	3.38	3.33	3.11	3.20	2.91
Other staple crops	2.44	2.68	2.46	3.18	2.92	2.77	3.23	2.76	1.97
Cotton	2.10	2.81	3.55	3.37	2.51	1.80	1.20	0.86	0.56
Oil crops	6.91	6.51	6.41	7.16	6.37	5.93	3.94	4.02	4.14
Sugar crops	1.07	1.59	1.34	0.85	0.95	0.82	0.90	0.67	0.43
Flax	0.05	0.01	0.00	0.00	0.01	0.02	0.05	0.01	0.01
Tobacco	0.98	0.89	0.62	0.63	0.63	0.65	0.55	0.46	0.43
Mulberry	1.27	1.38	1.30	1.04	0.72	0.53	0.33	0.45	0.34
Other cash crops	1.65	1.87	1.91	1.63	2.04	2.72	2.02	2.55	3.46
Vegetables	5.28	5.56	5.39	5.87	5.96	5.91	6.69	6.96	6.64
Fruits	8.03	6.66	8.35	9.15	8.59	8.41	8.69	9.14	9.43
Other horticultural crops	1.53	1.53	1.16	1.47	1.34	1.72	1.61	2.71	2.87
Agriculture as main Income source (%)	0.82	0.80	0.80	0.80	0.78	0.76	0.77	0.76	0.75
Non-crop income	12.02	13.72	17.19	17.29	18.69	21.31	24.42	25.22	27.98
Computer	0.01	0.01	0.01	0.02	0.04	0.07	0.09	0.14	0.16
Television	0.91	0.90	0.93	0.95	0.94	0.97	0.95	0.97	0.97
Fertilizer price (yuan/kg)	1.51	1.50	1.53	1.94	1.78	1.69	1.91	2.03	1.93
Pesticide price (yuan/kg)	23.28	24.67	25.58	26.21	26.92	27.11	27.15	28.34	26.66
Extension service dummy	0.23	0.20	0.15	0.14	0.19	0.16	0.20	0.16	0.11
Subsidy dummy	0.08	0.05	0.03	0.03	0.06	0.02	0.06	0.03	0.02
No. of Observations	3,943	3,909	3,936	3,959	3,996	3,813	3,689	3,847	3,477

Note: the unit of non-crop income is 1000 yuan. All price and income variable have been deflated to 2005 level.

Appendix B. Crop Structure of Villages with and without Certification, 2005 and 2013

	2005		2013	
Crop structure (%)	C-villages*	N-villages*	C-villages	N-villages
Wheat	22.90	15.23	15.07	14.22
Rice	33.79	18.33	35.30	15.08
Maize	16.98	24.20	15.34	35.04
Soybean	1.95	6.34	0.86	2.51
Potato	0.00	3.82	0.40	3.18
Other staple crops	2.63	2.42	0.95	2.08
Cotton	0.00	2.36	0.00	0.62
Oil crops	1.55	7.58	2.28	4.34
Sugar crops	0.23	1.18	0.08	0.47
Flax	0.02	0.06	0.00	0.01
Tobacco	0.27	1.07	0.00	0.48
Mulberry	0.95	1.31	0.16	0.36
Other cash crops	1.36	1.69	3.08	3.50
Vegetables	5.15	5.23	5.29	6.77
Fruits	11.38	7.61	19.28	8.35
Other horticultural crops	0.84	1.57	1.91	2.99

Note: Share of each type of crops in total sown area calculated from RFOP data.
*: C-villages are villages that adopted certified food production at some point, and N-villages are village that never adopted certified food production.

Chapter 3 Resource scarcity and cooperation

evidence from an irrigation district in western China

Abstract: This study examines the impact of long-term exposure to resource scarcity on farmers' cooperation. A historical irrigation water quota system in western China provides an opportunity to measure exogenous variation of water scarcity within an otherwise homogeneous region. We use the ratio of the arable land to the irrigation water quota of each village as our measure of water scarcity. Moreover, we use survey questions to measure collective actions in irrigation activities and use a public goods game to measure cooperation norms in rural communities. We find that irrigation water scarcity not only induces better irrigation management practices and outcomes, but also fosters a stronger cooperation norm.

This Chapter is based on:

Zihan Nie and Xiaojun Yang (2018). Resource scarcity and cooperation: evidence from an irrigation system in western China. Working Paper.

3.1 Introduction

The literature on the impact of resource scarcity on people's behavior is mixed. Resource scarcity can be related to more competition, more conflicts and less cooperation. Long-term exposure to resource scarcity can induce anti-social behavior and incite conflicts (Prediger, Vollan, and Herrmann 2014), while resource abundance tends to reduce conflicts and wars (Brunnschweiler and Bulte 2009). Short-term resource scarcity caused by negative climate and economic shocks can incite conflicts as well (Miguel, Satyanath, and Sergenti 2004, Burke, Hsiang, and Miguel 2015, Maystadt and Ecker 2014). In experimental settings, artificially created scarcity can also undermine cooperative behavior (Pfaff et al. 2015, Blanco, Lopez, and Villamayor-Tomas 2015, Gatiso, Vollan, and Nuppenau 2015).

Yet, resource scarcity may promote cooperation if people perceive incentives to use the resources efficiently in order to maximize the welfare of the group. In experimental settings, people have refrained, at least to some extent, from over-appropriating when facing increasing scarcity and potential depletion of resources (Oses-Eraso, Udina, and Viladrich-Grau 2008, Oses-Eraso and Viladrich-Grau 2007, Lindahl, Crepin, and Schill 2016). In field settings, several studies suggest that (perceived) resource scarcity is associated with better resource management (Araral 2009, Wang, Chen, and Araral 2016, Ito 2012, Brooks 2010).

Another strand of literature highlights the influence of behavioral experience on social norms. Bowles (1998) identifies "task performing effects": when people work together toward a common goal, this can foster cooperation. Empirically, Carpenter and Seki (2011) and Gneezy, Leibbrandt, and List (2016) find that differences in work organizations affect such preferences. For instance, fishermen who work closely together are more cooperative than those who operate more individually. Attanasio, Polania-Reyes, and Pellerano (2015) find that regular meetings can improve cooperation.

This study aims to link these strands of literature and investigate how resource scarcity affects cooperation in rural communities. Specifically, we focus on water scarcity, and examine the effect of water scarcity on cooperation in the context of a gravity irrigation system in western China. Following literature on irrigation management (e.g. Araral 2009, Bardhan 1993, 2000, Ito 2012, Wang, Chen, and Araral 2016), we expect to see a positive effect of irrigation water scarcity on collective irrigation management activities. More importantly, following the "task performance effect" logic that collective action fosters cooperative norms, we expect that communities with scarcer irrigation water have stronger norms of cooperation. We build a

simple model of irrigation agriculture with endogenous norm formation to illustrate the coevolution of collective irrigation management behavior and the norm of cooperation and to evaluate how water scarcity affects both. We form our hypothesis based on the comparative statics of the long-term equilibrium.

Instead of artificially manipulating scarcity in experimental settings (e.g., Pfaff et al. 2015), we measure water scarcity in the field, using a unique irrigation water quota system in western China. Under the quota system, the amount of water allocated to each village was based on the irrigated land areas self-reported by the village before the construction of the reservoir and canals in 1960s. Villages that reported larger irrigated land areas got more irrigation water, but also had to undertake more workload in the construction of the reservoir and canals. The tradeoff between the benefits (more irrigation water) and costs (more labor input in canal construction) resulted in differences in the relative size of reported irrigation areas compared with the actual arable land areas, which created variations in water scarcity across villages in later agricultural activities. We use these variations as our measure of real life water scarcity. In addition, we collected rich information about farmers' irrigation management activities, and used a lab-in-the-field public good experiment to measure the norm of cooperation.

We find that water scarcity improves irrigation management in terms of both the irrigation-related activities and their outcomes: people living in more water-scarce villages are more likely to coordinate in crop choices, more likely to keep local canals clean, and have higher self-reported quality of canals. More importantly, we find that the impact of water scarcity goes beyond irrigation-related activities. People in villages with a higher level of water scarcity also make significantly higher contributions in the public goods game. This result suggests that water scarcity also strengthen the norm of cooperation among rural communities.

This study contributes to three strands of literature. First, it contributes to the emerging literature on the impact of contextual factors on the formation of social preferences and norms. An emerging strand of studies has investigated how, for example, market institutions (Bowles 1998), political regimes (Brosig-Koch et al. 2011, Kuhn 2013, Heineck and Sussmuth 2013), production technologies (Leibbrandt, Gneezy, and List 2013, Alesina, Giuliano, and Nunn 2013, Talhelm et al. 2014) and workplace organization (Carpenter and Seki 2011, Gneezy, Leibbrandt, and List 2016) shape social preferences and norms. This study examines the role of resource scarcity in promoting the cooperative norm. While many studies manipulate resource scarcity using experimental methods (Pfaff et al. 2015, Blanco, Lopez, and Villamayor-Tomas 2015, Gatiso, Vollan, and Nuppenau 2015) or focus on negative shocks on

resource availability (Cassar, Healy, and Von Kessler 2017), few have examined the effect of relatively long-term resource scarcity on cooperation. Similar to our study, Sarangi, Jha, and Hazarika (2015) also measure the historical resource scarcity, but they mainly investigate how resource scarcity affects the culture of gender inequality. In a study with pastoralists in Namibia, Prediger, Vollan, and Herrmann (2014) find that resource scarcity increases anti-social behaviors and "identify long-term exposure to greater scarcity as source of individual conflict behavior". On the contrary, our study finds that greater water scarcity increases cooperation among individuals in an irrigation system. This finding suggests that, while long-term exposure to resource scarcity can invoke potential conflicts, it is not always the case. In different settings (in our case, irrigation water in a gravity irrigation system), resource scarcity could foster and improve cooperation among individual users.

Second, this study contributes to the literature regarding the impact of the gravity irrigation system on cooperation. As it requires large scale investment and collective efforts to build, maintain and use a gravity irrigation system, irrigation per se is an institutional arrangement that could have a profound impact on people' preferences and social norms and thus on their societies. Tsusaka et al. (2015) find that the introduction of gravity irrigation increases farmers' altruism and cooperation in a short period of time. von Carnap (2017) identifies irrigation agricultural practices as a determinant of social capital in India and finds that different types of irrigation affect social capital formation differently. Bentzen, Kaarsen, and Wingender (2017) find that irrigation makes societies less likely to be ruled under democracy at the subnational level; people in areas with higher irrigation potential hold lower opinions of democracy, as the large investment required to build an irrigation system favors elite monopolization of water and thus political power. Our study shows that such impacts are not limited to comparisons between people or society with and without irrigation: within a gravity irrigation system, the intensity of the irrigation institution (through collective irrigation management due to water scarcity) also affects the norm of cooperation among farmers.

Lastly, our research also speaks to the literature on common pool resource (CPR) management. Several studies have linked CPR management with public goods provision (Botelho et al. 2015, Solstad and Brekke 2011), where they find that allowing resource users to contribute to a public good with their gains from the resource could help to mitigate the over-extraction problems. Our study provides a different link between resource extraction and public goods, where the effort of appropriating resources itself is a public good. If resource appropriation requires collective actions, as in the case of gravity irrigation, it could help foster

a norm of cooperation, which could spill over to practices in environmental and resource management.

The rest of this paper is organized as follows: Section 3.2 provides a simple model to illustrate how water scarcity could affect the norm of cooperation; Section 3.3 describes our empirical strategy and the data; Section 3.4 presents empirical results; and Section 3.5 concludes the paper.

3.2 A model with endogenous norm formation

In this section, we build a simple model of endogenous norm formation in the context of a gravity irrigation system to demonstrate how water scarcity affects the evolution of the norm of cooperation within rural communities through the demand for collective actions.

Assume a rural community where agricultural production relies on irrigation and the irrigation water (W) available to the community is exogenously determined.[19] For simplicity, we assume that the agricultural output q_i is a C-D production function of two inputs, private labor input (l_i) and irrigation service (A): $q_i = {l_i}^{\alpha} A^{\beta}$, $\alpha, \beta \in (0,1)$.[20] The irrigation service A can be interpreted as irrigation water use efficiency, which is a public good that is shared by every member of the community. We assume that water use efficiency $A = A(G, W)$ is a function of irrigation water (W) and the collective efforts on irrigation management (G) from the community on irrigation-related activities such as maintenance and coordination in farming decisions. For a community with exogenously given W, A is determined by the collective effort G and thus we can consider G as a public good as well. Collective effort $G = \sum_{i=1}^{N} g_i$ is the sum of the individual effort on irrigation management of all N households in the community. It is reasonable to assume that A is concave to G and W: $A_1 = \frac{\partial A}{\partial G} > 0$, $A_2 = \frac{\partial A}{\partial W} > 0$, $A_{11} = \frac{\partial^2 A}{\partial G^2} \leq 0$ and $A_{22} = \frac{\partial^2 A}{\partial W^2} \leq 0$. Furthermore, we assume that, at least when the total amount of irrigation water falls in a certain range $W \in (\underline{W}, \overline{W})$, where water is moderately scarce, the marginal efficiency gain from additional collective effort is decreasing with W: $A_{12} =$

[19] Here, water supply W can be interpreted as the water supply per unit of land, which reflects the degree of water scarcity. As we assume that every household in the community has the same area of land, it reflects water scarcity on both household and communities level.

[20] The C-D production function form implies that a certain level of collection effort is optimal for each individual farmer despite free-riding incentives. This assumption suits the case of gravity irrigation, as something has to be done if anyone wants to get any water.

$\frac{\partial^2 A}{\partial G \partial W} < 0$, as implied in Ito (2012). Individual farmers have to decide how to divide their fixed labor endowment (E) into private farming (l_i) and irrigation service (g_i): $l_i + g_i = E$.

We further assume that farmers not only gain utility from material outputs from agricultural production, but also care about (complying with) a social norm regarding public goods provision. This social norm defines the appropriate contributions to public goods in the community. At any moment, the norm is acknowledged by everyone and applies to everyone. Failing to conform to this norm would result in costs such as reputation loss or psychological discomfort. Contributions that exceed the norm may also benefit people (e.g., by building better reputation). We assume the utility from (not) conforming to the norm takes the form of $R(\Delta g_i, r_i) = r_i f(\Delta g_i)$, which depends on the gap ($\Delta g_i = g_i - g^N$) between individual contribution (g_i) and the norm (g^N) and individuals' sensitivity to the norm (r_i), where $r_i \in [0, r^{max}]$. We further assume that $R(\Delta g_i, r_i)$ has the following properties: $R(0, r_i) = 0$, $f' > 0$, $f'' < 0$. These properties mean there is no utility gain or loss when people exactly match the norm; there is utility gain when own contributions exceed the norm and utility loss when own contributions fall short of the norm; and the marginal utility loss from falling short of the norm is larger than the marginal utility gain from exceeding the norm.

For simplicity, we assume farmers have a linear utility function $u_i = q_i + R_i$, where q_i is agricultural output and R_i is the utility related to the social norm. Then, farmer i faces the following maximization problem:

$$\max_{l_i, g_i} u_i = {l_i}^{\alpha} A^{\beta} + r_i f(\Delta g_i),,$$

$$\text{s.t. } l_i \geq 0,\ g_i \geq 0 \text{ and } l_i + g_i = E.$$

Proposition 1: In the short term, when the social norm on public goods provision g^N is given, individual's optimal public goods provision $\widetilde{g_i}$ is increasing with social norm g^N, increasing with sensitivity to social norm r_i, and decreasing with water resource W.

Proof: see Appendix.

So far, we treat the norm of appropriate public goods contribution g^N as given for individual decision making. But social norms also evolve. For simplicity, we assume that norms evolve in a "naive" way, in that people in a community form and adjust the norm of appropriate public goods provision according to the average contribution observed in the previous period: $g^N_{t+1} = \int_0^{r^{max}} \widetilde{g_{i,t}}\, h(r_i) dr_i$. $h(r_i)$ is the probability density function of r_i.

Proposition 2: The long-term equilibrium level of social norm $\widehat{g_t^N}$ is decreasing with water resource *W*.

Proof: see Appendix.

Thus, in the long term, people living in villages with a higher level of water scarcity have better cooperation in public goods provision and higher social norms regarding public goods provision. How do these predictions translate into behavior in the abstract setting of a linear public goods game?

Proposition 3: If people follow decision-making heuristics (Tversky and Kahneman 1974) or a rule of thumb that carries the established norms to other situations, such as a public goods game, given individual sensitivity to the norm r_i, we should observe that people in villages with a higher level of water scarcity (smaller *W*) contribute more to the public goods than people in villages with lower level of water scarcity.

Proof: see Appendix.

The simple model above illustrates how water scarcity affects cooperation in public goods provision through endogenous norm formation. A higher level of water scarcity increases the marginal payoff contribution to irrigation service, which in turn fosters a cooperative norm. Then, guided by the more cooperative norm, people tend to have higher contribution in settings beyond agricultural and irrigation activities, such as in a public goods game. At the long-term equilibrium, the predictions of the model give us testable hypotheses about the relationships among water scarcity, irrigation management, and the norm of cooperation. Water scarcity is expected to improve both irrigation management and the norm of cooperation in rural communities.

3.3 Empirical strategy and data

3.3.1 Study site: irrigation water quota system in Minle

We carried out our study in Hongshui River Irrigation District in Minle County, Zhangye City, Gansu Province in northwestern China. Minle County is an oasis located in the northern foothills of the Qilian Mountains and lies in the middle of the Hexi corridor, which is characterized by a semi-arid climate and a long history of irrigation agriculture.[21] Agriculture

[21] Zhang et al. (2013) provide a detailed introduction to the socio-economic and geo-climatic background of Minle County.

in Minle depends heavily on irrigation and water is the main constraint on agricultural production. The traditional and main source of irrigation water is surface water from local rivers. Hongshui River is the largest of the five major rivers in Minle County. Hongshui River Irrigation District is the largest irrigation district in the county and one of the "large irrigation districts" at the national level. We choose to focus on only one irrigation district in one county in order to eliminate potential confounding factors such as different geo-climate conditions, different socio-economic histories, different irrigation cultures and traditions, and different local regulations and policies.

The distinctive feature of the irrigation practices in this area is the water quota system. The Hongshui River Irrigation District was the first to adopt the water quota system. The irrigation water quota system in Hongshui River Irrigation District was first introduced in 1966 when facing the construction of the new irrigation canals and the reservoir. Villages that would benefit from the new canal system were asked to report their irrigated land areas. These self-reported irrigated land areas were then used as the sole criterion for the allocation of both irrigation water and the workload of irrigation infrastructure construction and maintenance across villages. Villages that reported larger irrigated land areas would receive more irrigation water but also had to undertake a larger obligation in the irrigation infrastructure construction and maintenance. These self-reported irrigated areas became the measure of water quota and are called "*pan ding pei shui mian ji*", or "determined water allocation areas". We will refer to these self-reported irrigated areas as "irrigation water quota" or "water quota" in the rest of the paper.

Because the availability of irrigation water was tied to labor input obligations, when reporting the irrigated area, villages had to balance between the benefits (more irrigation water) and costs (more labor input in canal construction rather than on their own land). Such trade-offs resulted in differences between irrigation water quota and the actual land size across villages. Villages that received relatively more water quota suffered less from water scarcity in later years and *vice versa*, which created variations in the relative water scarcity across villages. In this study, we take advantage of these historically formed variations in water scarcity to examine the impact of water scarcity on cooperation.

3.3.2 Measuring water scarcity

Measuring water scarcity is crucial to our empirical analysis. The level of scarcity depends on the demand for water as compared to the supply of water. Therefore, we define our water scarcity indicator based on the ratio of the potential demand for water to the accessible irrigation water supply of each village:

$$Scarcity = \frac{Village\ arable\ land\ areas}{Village\ irrigation\ water\ quota}.$$

The irrigation water quota of each village, shown in the denominator, determines the supply of irrigation water, while arable land area in each village, shown in the numerator, reflects the potential demand for water.[22] The ratio represents how much arable land one unit of water quota has to irrigate.[23] Water is scarcer when the ratio is larger. We adjust the water quota and land size to any changes that have occurred since the land reform to capture the initial state of water scarcity.[24] We believe that this ratio can largely capture the differences in water scarcity between villages over a relatively long time.

Note that our measure of water scarcity only refers to surface water, while groundwater availability could affect the validity of our scarcity measure. However, the groundwater is not an important issue for two reasons. First, while we focus on long-term exposure to water scarcity in this study, the use of groundwater is a recent phenomenon in Minle. The oldest well in our sample villages was dug in 1987, the second oldest one was dug in 1998, and there were only five wells before 2009. The boom of well digging occurred in 2011 and 2012, and 58% of the wells were dug in these two years.[25] Access to groundwater does not affect the initial differences in water scarcity; if anything, the use of groundwater could be the result of initial

[22] We use the potential demand for water (represented by arable land of each village) instead of actual demand measures such as sown area and crop portfolio, because the actual demand is likely to be endogenous to the availability of irrigation water. Using arable land size as the numerator captures long-term scarcity, which is the focus of this study, better than using actual water demand measures in 2015.

[23] The ratio is used to compare the level of water scarcity among villages within Hongshui River Irrigation District. It is not an indicator that is comparable to water stress level in other regions. When comparing Minle with the rest of Gansu or China, we believe the extent of irrigation water scarcity can be seen as "moderate". Water is scarce in the sense that water is the single most binding constraint to local agricultural production. However, we regard water in Minle as not being very scarce because irrigation agriculture in Minle is still very active, and Minle and neighboring areas have been famous for their irrigation agriculture throughout history.

[24] Changes in the water quota have been rare. Only six out of the twenty-six villages in our sample reported that they had a change in village water quota, and these changes were small. The most recent change was in 1997. There have been no changes in the arable land areas of our sampled villages since the land reform in 1981. This is probably because our study site is the central part of the irrigation district and has been fully reclaimed. Expansion of arable land happened mostly in marginal areas of the irrigation district. We do not mean to extrapolate this trend outside of our sample.

[25] There are 90 wells in total in the 26 villages, but we only have the information on which year they were dug for 79 wells. Only two villages reported when their first wells were dug (in 2002 and 2011).

water scarcity. Second, despite the increasing access to groundwater in recent years, groundwater is still not an important source for agricultural irrigation (Zhang et al. 2013). The cost of pumping groundwater out of the deep wells is much higher than that of surface water. Most importantly, groundwater salinity is much higher than surface water salinity, making groundwater harmful to yields and soil fertility. Therefore, groundwater is only used as a complement to surface water. Irrigation water scarcity is mainly driven by access to surface water.

Our water scarcity measure based on the irrigation water quota system has several merits. First, shocks and uncertainty in water supply affect all villages simultaneously in the same direction. In the water quota system, when there is more (less) water available, each water quota unit gets more (less) water, and thus the relative scarcity across villages, represented by the size of the scarcity ratio, is not affected. Therefore, although literature have shown that uncertainty in the size of resources could affect people's behavior (e.g. Aflaki 2013), our measure of water scarcity is not confounded by the variations and uncertainty in water supply. Second, because the quota system sets the maximum amount of water that a village can use, the water use of one village within its quota does not affect how much water other villages can use.[26] Because the water quota of each village and the time of irrigation are public knowledge in the whole irrigation district, and because irrigation can be easily controlled by the sluices on the main canal, the chances of stealing water from other villages are very low. Inter-village competition over water is very unlikely and thus cannot affect people' preferences and norms formed within their villages. Third, because the amount of water that each quota unit represents is set by the local irrigation district administration, water use today does not affect the stock of water resource and hence the water flow tomorrow, provided that the irrigation district administration allocates water reasonably. Therefore, the typical intertemporal tradeoff between today's and future's consumption, the dynamics between resource stock and flow, and the externality of individuals' extraction in a common pool resource scenario, play little role in determining famers' behaviors in our case. This helps rule out the endogeneity of resource scarcity due to past extraction. These features greatly simplify our analysis and enable us to focus on the impact of water scarcity itself.

[26] Traditionally, water quotas were not transferable. The market for water quota reform was only established after a water rights reform pilot project, Building a Water-saving Society in Zhangye in 2002. Yet, water right transactions have been rare despite the existence of the market. (Zhang et al. 2009) This study focuses on the effect of historical water scarcity. This recent event should not affect our analysis. If anything, recent water right transactions could be the consequence of historical water scarcity.

It is crucial that the variations in our water scarcity measure are exogenous to people's social preferences and social norms of cooperation. There are three important aspects that support our claim. First, the water quota of each village was not decided by the people we interviewed, but by the older generation. Even the oldest person in our sample (age 66) was a teenager in 1966 when the water quota was determined. They might have been involved in the later construction of the canals and the reservoir, but they were certainly not involved in deciding how much irrigated land should be reported to the county authority. We can safely say that the irrigation water quota and hence the degree of water scarcity was imposed upon them, rather than determined by them. Second, migration across villages has been rare. Cross-village migration could affect our identification strategy in two ways. First, cross-village migrants would have experienced different levels of water scarcity, which would jeopardize our water scarcity ratio as a measure of exposure to water scarcity. Second, if people migrate among villages as a response to water scarcity, it raises a potential selection bias issue. However, the Chinese *hukou* system[27] largely ties people to where they were born, especially for rural residents. Therefore, migration due to water scarcity was unlikely. Marriage has been a legitimate reason for inter-village migration, but it is much more common for women than for men. In our all-male sample, only four out of the 312 subjects were not born in the village in which they lived at the time of the survey. Therefore, cross-village migration is not an important issue for our study. Third, unlike in many common pool resource settings where a more cooperative community could better preserve its resource stock and alleviate resource scarcity, the water quota system limits this channel of endogeneity, because current water flow is not directly affected by individuals' or communities' water use in the past.

However, there is still one potential source of endogeneity that we cannot rule out: the intergenerational transmission of social preferences or norms. If the older generations were more pro-social and thus were willing to contribute more to the public projects in exchange for more water, or if they formed a more cooperative culture or social norm through working together on the public projects, their pro-social preferences or cooperative norms could be transmitted to the younger generations. Then, either the current water scarcity in terms of irrigation water quota is the result of certain social preferences, or both the scarcity and the social preferences are results of some omitted factors. Yet, as we will discuss later, even if such endogeneity exists, it does not jeopardize our main findings.

[27] *Hukou* is the household registration system in China.

3.3.3 Measuring irrigation management

The most straightforward measure of effort in irrigation management activities is the actual labor or monetary contribution to irrigation-related collective activities. For instance, Ito (2012) uses the household labor contribution to irrigation management as the measure of collective actions of farmers. However, such labor or monetary contributions are quite noisy, because current contributions could be affected by the contributions in the past. For example, a large investment in upgrading a canal system in the past could lead to less maintenance effort now. Unfortunately, we only have one-year cross-sectional data with a relatively small sample size, which is not enough to smooth such noise. Alternatively, we turn to irrigation-related activities that are carried out more frequently and bear fewer long-term implications. Specifically, as summarized in Table 1, we mainly focus on two aspects: coordination in farming and keeping the canal clean. Farming coordination can improve irrigation efficiency because the same type of crops planted closely together can be irrigated at one time when water is most needed, which thus reduces the water loss from multiple rounds of irrigation. Coordination in farming includes two dummy variables: whether a farmer coordinates with other farmers about crop portfolios and whether a farmer discusses what to grow with farmers who have neighboring plots. Cleaning canals is a major canal maintenance work every year and keeping the canals clean can help ease the job for everyone and improve irrigation efficiency. Keeping the canals clean includes three dummy variables: whether people dump trash in the canal, whether they stop other villagers from doing so, and whether they stop strangers from dumping trash in the canal. In addition, we expect that better irrigation management is reflected in its outcomes. We use the self-reported canal quality to measure the outcomes of collective irrigation management. The indicator is a five-point scale self-evaluation on the quality of the third tier, fourth tier and fifth tier canals.[28] If water scarcity motivates farmers to invest more in maintaining and renovating canals, such efforts are expected to result in higher canal quality. We expect that a higher level of water scarcity leads to better irrigation canal quality.

[28] The irrigation canal network in the Hongshui River District typically has five tiers. The first-tier canal is the main canal that connects directly to the reservoir. The second-tier canals are the branches of the first-tier canal and carry irrigation water to multiple villages. Both of these higher-level canals are governed by higher levels of administration and are not directly involved in village level irrigation. Thus, we didn't include them in the questions on canal quality. From the third-tier down, canals are more directly involved in village irrigation. The third, fourth and fifth tier canals together are often referred as the "end level canal network (*Mo Ji Qu Xi*)". They are the canals that directly irrigate the fields and are often shared and maintained by a group of farmers.

Table 1. Questions and indicators of farmers' effort on irrigation-related collective activities

Variable	Survey question	Variable define	Mean (SD)	N
A: Coordination in farming				
Crop structure decide	How did you decide what crops to grow in 2015?	Dummy variable: 0, decide on their own; 1, decide through coordination.	0.42 (0.49)	297
Crop discuss	Do you discuss with neighboring farmers about which crop to grow every year before sowing?	Dummy variable: 1, discuss; 0, don't discuss.	0.83 (0.38)	311
B: Keeping the canal clean				
Throw trash	Do people dump trash in the canal?	Dummy variable converted: 1, never or rarely; 0, often;	0.73 (0.45)	310
Stop villager	If you see other villagers damp trash in the canal, will you stop him?	Dummy variable, converted from 1-5 scale: 1, probably or definitely will stop him; 0 otherwise.	0.91 (0.29)	312
Stop stranger	If you see a stranger damp trash in the canal, will you stop him?	Dummy variable, converted from 1-5 scale: 1, probably or definitely will stop him; 0 otherwise.	0.95 (0.21)	312
C: Self-reported canal quality				
Canal quality: third tier	How is the condition of the third-tier canal?	1-5 scale, 1 as the worst condition and 5 as the best	3.38 (1.37)	281
Canal quality: fourth tier	How is the condition of the fourth-tier canal?	1-5 scale, 1 as the worst condition and 5 as the best	2.50 (1.26)	216
Canal quality: fifth tier	How is the condition of the fifth-tier canal?	1-5 scale, 1 as the worst condition and 5 as the best	1.89 (1.23)	219

Note: calculated by the authors using the household survey data.

3.3.4 Measuring cooperation: a public goods game

We conduct a lab-in-the-field experiment to measure the more general inclination of cooperation beyond agriculture- or irrigation related activities. Specifically, we use a repeated linear public goods game to measure cooperation.[29] In the abstract and anonymous setting of a public goods game, there is no incentive for a self-interested person to contribute. Thus, contribution behavior in a public goods game can be used as a measure of the norm of cooperation in rural communities. We briefly summarize the game design here.

Subjects are randomly divided into three-person groups at the beginning of each round. Each subject receives an initial endowment of 10 *Yuan* in each round.[30] They are asked to

[29] This standard public goods game is part of a three-stage public goods game design. After this stage of the five-round public goods game, subjects were asked about their preferences between a punishment and a reward institution. They then were randomly assigned to one of the institutions for another five rounds of a public goods game with punishment or reward. In addition, subjects also played a simple risk game and three binary-choice dictator games. The whole experiment lasted 60-90 minutes, and the average payment to subjects was 166 *Yuan*.
[30] At the time of the experiment, 1 USD=6.58 CNY.

decide how much to keep in a personal account and how much to contribute to a group account. The payoff function is described as follow:

$$\pi_i = 10 - g_i + 0.5 \sum_{j=1}^{3} g_j,$$

where π_i is subject i's payoff and g_i is his contribution to the group account. The amount of the group account equals the sum of the contributions from the three subjects in the same group. The marginal payoff of the group account is 0.5, offering monetary incentives to free ride. After subjects made their contribution decisions, they were informed about the total group contribution and their individual payoffs before proceeding to the next round. The public goods game lasted for five rounds. Subjects were randomly distributed into three-person groups at the beginning of each round.

We use the subjects' contributions to the group account as our measure of cooperation in each village. Specifically, we use both the average contributions over the five rounds and the contributions in the first round as our key dependent variables.

We conducted the experiment in January 2016 along with a household survey. The experiment was computerized and programmed with z-Tree (Fischbacher 2007). We created a lab environment with tablets and cardboard boxes in the conference rooms of village administration buildings. We turned cardboard boxes into small cubicles with tablets inside so that subjects could make their decisions independently and anonymously. Communication among subjects was not allowed. Subjects received oral instructions from the experimenters at the beginning of the experiment and were asked to answer practice questions on paper. The practice questions aimed to test whether the participants understood how to calculate the payoff from the contributions. The experiment only proceeded when all subjects were able to correctly answer the practice questions. In addition to the experiment, all subjects also participated in a household survey.

3.3.5 Subject Pool

We randomly selected 26 villages in the Hongshui River Irrigation District in Minle County and selected 12 men from each village as our experimental subjects and survey respondents. We chose male-only subjects for several reasons. Men are usually household

decision makers and represent the family in most community events.[31] More important, compared to women, men are more exposed to irrigation activities and water scarcity than are women, because men are traditionally more involved in agricultural production and irrigation-related activities. Men are also much less likely than women to migrate across villages, so that issues related to cross-village migration are avoided. Moreover, most middle-aged women in Minle Country have little education and had trouble understanding the setup of the experiment in our pilots.[32] Therefore, we limited the subjects to males.

Our target age range was from 40 to 65 years old. We set this age range because we want to target the people who have engaged in irrigation activities and been exposed to irrigation water scarcity for a relatively long period. We exclude older men because they were adults when the water quota system was established and they might have been influenced by the experience of canal and reservoir construction; they might even have played a role in determining water quota. We exclude younger generations because they have been less exposed to water scarcity and less engaged in agricultural activities.

The summary statistics of individual and household characteristics of the subjects are presented in Table 2. The vast majority of our subjects are household heads. The average age is 51 years old.[33] The average educational level is quite low (primary school education). All subjects are ethnic Han. Only two subjects reported having urban hukou, while the rest have rural hukou. The subjects on average spent more than 8 months in the village in 2015. Thirty-nine percent of the subjects were employed in off-farm jobs in 2015. The average gross household income per capita is 22,060 *Yuan*. For the village characteristics, 35 percent of the sampled villages have non-farming enterprises and the average share of local off-farm labor is 19 percent. As the key variable for our study, we can see that the water scarcity ratio is 1.38 with moderate variations (ranging from 1 to 2.02; coefficient of variation is 20%).

[31] It has been found that women on average have lower bargaining power than their husbands in the context of rural China (Bulte, Tu, and List 2015, Yang and Carlsson 2016).

[32] Thirty-nine percent of the wives of our subjects never received any formal school education and only seven percent received education higher than elementary school, while the corresponding figures for our male subjects are three percent and fifty-seven percent.

[33] Although we targeted people from 40-65, in the process of survey implementation, there were actually two men younger than 40 years old (35 and 37 years old) and two men older than 65 years old (66 years old) included in our sample. But we don't believe this is a serious issue, and this does not affect our main results.

Table 2. Summary statistics of key variables

Variable name	Mean	SD	Min	Max	N
Individual characteristics					
Household head (dummy)	0.97	0.16	0	1	312
Age	51.4	4.96	35	66	312
Years of schooling	6.63	2.62	0	15	312
Ethnic dummy (han=1)	1	0	1	1	312
Hukou dummy (rural=1)	0.99	0.08	0	1	312
Off-farm job dummy (have any=1)	0.39	0.49	0	1	312
Time at home in 2015 (month)	8.24	3.45	1	12	312
Household characteristics					
No. of siblings	4.07	1.97	0	11	312
Majority Family name dummy (yes=1)	0.83	0.38	0	1	312
Household size (person)	4.09	1.35	1	8	312
Farm size (mu)	18.36	10.94	2.5	72	312
Gross income per capita (1000 yuan)	22.06	56.23	0.06	753.41	312
Village characteristics					
Village arable land size (mu)	4959	2265	1500	11200	312
Distance to town seat (km)	5.17	2.58	1	11	312
Distance to county seat (km)	13.79	99.51	0	40	312
Village enterprise dummy	0.35	0.48	0	1	312
Share of local nonfarm labor (%)	19.01	11.54	0.75	50.61	312
Village water quota (mu)	3807	1708	1064	9520	312
Water scarcity (ratio)	1.38	0.28	1	2.02	312

Note: calculated by the authors using the household survey data.

3.3.6 Empirical models

To empirically test our hypothesis, the model in Section 2 implies a dynamic structural model. However, because we only have one-year cross-sectional data and we are mainly interested in the long-term impacts of resource scarcity, not the dynamic relationship between irrigation management activities and the social norm, we estimate the impact of water scarcity on irrigation management and contributions in the public goods game (PGG) separately using the following reduced-form models:

$$IrriM_{ij} = \alpha_0 + \alpha_1 Scarcity_j + \alpha_2 X_{ij} + \alpha_3 V_j + \epsilon_{ij};$$

$$Coop_{ij} = \beta_0 + \beta_1 Scarcity_j + \beta_2 X_{ij} + \beta_3 V_j + \varepsilon_{ij},$$

where $IrriM_{ij}$ stands for irrigation management indicators and $Coop_{ij}$ represents the contributions in PGG of subject *i* in village *j*; $Scarcity_j$ is the water scarcity indicator for village *j*; X_{ij} is individual and household characteristics of subject *i* in village *j*; V_j represents characteristics of village *j*; and ϵ_{ij} and ε_{ij} are the village-clustered error terms. Specifically, individual and household characteristics *X* include age, years of schooling, contracted land size, number of siblings, household size and a dummy variable for having an apartment in town as an indicator for wealth. Village characteristics *V* include logarithmic arable land size, distance to town seat, distance to county seat, dummy for having non-farming enterprises in the village

and percentage of off-farm labor. We choose these explanatory variables based on literature that identifies factors that could affect collective actions (Araral 2009, Wang, Chen, and Araral 2016).

The coefficients of water scarcity indicators, α_1 and β_1, are the key parameters in which we are interested. Our model predicts that the level of water scarcity is positively associated with both irrigation management and contributions in the public goods game. Thus, we expect to find positive α_1 and β_1 in all model specifications. We vary the estimation methods according to the different data structures of dependent variables, and the detailed estimation strategies will be discussed in the results section.

3.4 Results

3.4.1 Water scarcity and irrigation management

First, we test whether the higher level of water scarcity leads to better cooperation in irrigation management. Table 3 displays the probit regressions results of the impact of water scarcity on farming coordination and canal maintenance. Consistent with our expectation, the results show that the coefficients of water scarcity indicators all have positive signs, although only three of them are statistically significant. We also build an index variable by adding up these five dummy variables and the OLS regression results are shown in the last column of Table 3. The results are consistent with the Probit results. Water scarcity induces better irrigation management.

Better irrigation management should also be reflected in its outcomes. Therefore, we further look at whether villages with a higher level of water scarcity have higher quality canals. We report the OLS regression results in Table 4.[34] We find positive effects of water scarcity on the quality of fourth- and fifth-tier canals. As the fourth-tier canals and fifth-tier canals are shared and managed by local farmers as local public goods, these results are consistent support for our previous argument.[35] Yet, we do not find a significant effect on the third-tier canal quality. This is probably due to the fact that the investment to improve third-tier canal quality

[34] Given that the dependent variables are on a 1-5 scale, we also run ordered an probit regression. The results are similar. For simplicity of interpretation, we only report OLS results here.

[35] Almost all fourth- and fifth-tier canals are shared by farmers and they maintain these canals together. In our survey data, 210 out of 215 households reported sharing fourth-tier canals with others and maintaining the canals together; similarly, 205 out of 220 households reported sharing fifth-tier canals and 203 out of 221 households reported maintaining fifth-tier canals together with others.

is quite large, so that it is often beyond the reach of the villages but has to rely more on investment from higher authorities. To sum up, we do find that water scarcity increases farmers' efforts in collective irrigation management, in both irrigation management activities and their outcomes.

Table 3. Water scarcity and irrigation management activities

	(1) Crop structure decide	(2) Crop discuss	(3) Throw trash	(4) Stop villager	(5) Stop stranger	(6) Irri management index
Water scarcity	0.183*	0.0497	0.234**	0.0364	0.124**	0.621***
	(0.107)	(0.0895)	(0.0950)	(0.0839)	(0.0559)	(0.181)
Age	-0.00171	-0.00673*	0.000774	0.00262	0.00110	-0.00420
	(0.00514)	(0.00366)	(0.00458)	(0.00328)	(0.00238)	(0.0102)
Years of schooling	-0.00465	0.00109	-0.00663	0.0118**	-0.00556	-0.00175
	(0.0114)	(0.0104)	(0.00965)	(0.00514)	(0.00579)	(0.0267)
Contracted land (mu)	-0.00381	-0.000665	-0.00427	-0.000343	-0.000813	-0.0108**
	(0.00273)	(0.00185)	(0.00304)	(0.00140)	(0.00109)	(0.00449)
No. of siblings	0.0112	0.0165	-0.00787	0.00994	0.00397	0.0329
	(0.0144)	(0.0114)	(0.0154)	(0.00705)	(0.00594)	(0.0357)
Household size	-0.00339	0.0223	0.00396	0.0266**	0.00175	0.0520
	(0.0266)	(0.0151)	(0.0151)	(0.0112)	(0.00587)	(0.0378)
Wealth: Have an apartment	0.135	0.222*	-0.0438	0.0579	-0.00813	0.339**
	(0.0897)	(0.114)	(0.0839)	(0.0615)	(0.0372)	(0.129)
ln(village arable land)	0.0260	0.0820*	0.00762	-0.0298	0.00843	0.137
	(0.0656)	(0.0475)	(0.0750)	(0.0505)	(0.0254)	(0.128)
Village enterprise dummy	0.0596	-0.0238	-0.114**	-0.0277	0.00168	-0.139
	(0.0529)	(0.0463)	(0.0567)	(0.0375)	(0.0225)	(0.128)
Share of off-farm labor in village	0.576**	0.358*	-0.253	-0.132	-0.177*	0.428
	(0.226)	(0.186)	(0.181)	(0.125)	(0.0980)	(0.366)
Distance to town seat	0.0393***	-0.00140	0.00130	0.00800	0.00658	0.0554**
	(0.00858)	(0.00927)	(0.00814)	(0.00728)	(0.00499)	(0.0214)
Distance to county seat	0.000395	0.00107	-0.00262	0.000586	0.000224	-0.000939
	(0.00277)	(0.00284)	(0.00298)	(0.00204)	(0.00125)	(0.00758)
Observations	297	311	310	312	312	294
Pseudo R^2 (adjusted R^2)	0.074	0.075	0.038	0.077	0.086	0.052

Note: Probit regression of behavior in and attitude toward irrigation related activities in column (1) – (5) and OLS results in column (6). Dependent variables in column (1) – (5) are dummy variables: *crop decide* equals 1 if there is coordination in determine crop structure; *crop discuss*, 1 as discuss with neighbors about what to grow on which land parcel; *throw trash*, 1 if one reports people never or rarely throw trash in the irrigation canal; *stop village*, 1 if he claims he would probably or definitely stop a villager from throwing trash in the canal; *stop village*, 1 if he claims he would probably or definitely stop a stranger from throwing trash in the canal. The dependent variable in column (6) is an index variable that equals the sum of the dummies variables used in column (1) – (5). In column (1) – (5), the marginal effects are reported. Robust standard errors in the parentheses are adjusted for clustering at village level. ***, ** and * stand for significant level 1%, 5% and 10%, respectively.

Table 4. Water scarcity and self-reported canal quality

	(1) Canal quality: Third tier	(2) Canal quality: Fourth tier	(3) Canal quality: Fifth tier
Water scarcity	0.288	0.950*	0.715*
	(0.462)	(0.479)	(0.372)
Age	-0.0165	-0.0168	0.0248
	(0.0146)	(0.0159)	(0.0204)
Years of schooling	-0.00447	-0.0313	-0.0337
	(0.0443)	(0.0349)	(0.0294)
Contracted land (mu)	-0.00545	0.0239**	0.0166
	(0.0120)	(0.0113)	(0.0143)
No. of siblings	0.00737	0.00238	-0.0347
	(0.0378)	(0.0387)	(0.0435)
Household size	-0.0276	-0.0687	-0.0125
	(0.0633)	(0.0690)	(0.0731)
Wealth: Have an apartment	0.309	0.574*	0.142
	(0.272)	(0.286)	(0.230)
ln(village arable land)	-0.152	-0.549	-0.0782
	(0.445)	(0.377)	(0.230)
Village Average income per capita	-0.338	0.181	0.102
	(0.368)	(0.349)	(0.224)
Dummy for enterprise in village	-0.985	-1.358	-0.360
	(0.894)	(1.088)	(0.734)
Share of non-farm labor in village	0.0376	0.0980**	0.0345
	(0.0405)	(0.0439)	(0.0383)
Distance to town seat	-0.00235	0.0345*	0.0102
	(0.0154)	(0.0177)	(0.0137)
Distance to county seat	5.435	5.956*	0.107
	(3.656)	(3.105)	(2.319)
Observations	281	216	219
Adjusted R2	-0.008	0.114	0.036

Note: OLS results for self-reported canal quality. Canal quality variables are 1-5 scale variables, where 1 stands for the worst quality and 5 is the best quality. We asked each villager about the quality of the canals that irrigate their lands. Not all household use all three tiers of canals to irrigate their lands, which leads to a large number of missing value in the sample, especially for the fourth and fifth tier canals. Robust standard errors in the parentheses are adjusted for clusters at village level. ***, ** and * stand for significant level 1%, 5% and 10%, respectively.

3.4.2 Water scarcity and contributions in the public goods game

So far, we have shown that water scarcity improves irrigation management activities and their outcomes in terms of canal quality. This is consistent with existing literature on the determinants of collective action in irrigation. But we are more interested in whether the effect of water scarcity goes beyond agriculture- or irrigation-related activities and shapes the norms of cooperation within rural communities. As predicted by our model following related literature such as Gneezy, Leibbrandt, and List (2016), we expect to find a positive effect of water scarcity on cooperation, i.e., higher contributions in the public goods game.

We first show the relationship between the water scarcity indicator and the contributions, in Figure 1A and 1B. Despite large variations in individual contribution at different degrees of

water scarcity, we can see that water scarcity and cooperation are positively correlated (the fitted line in red).

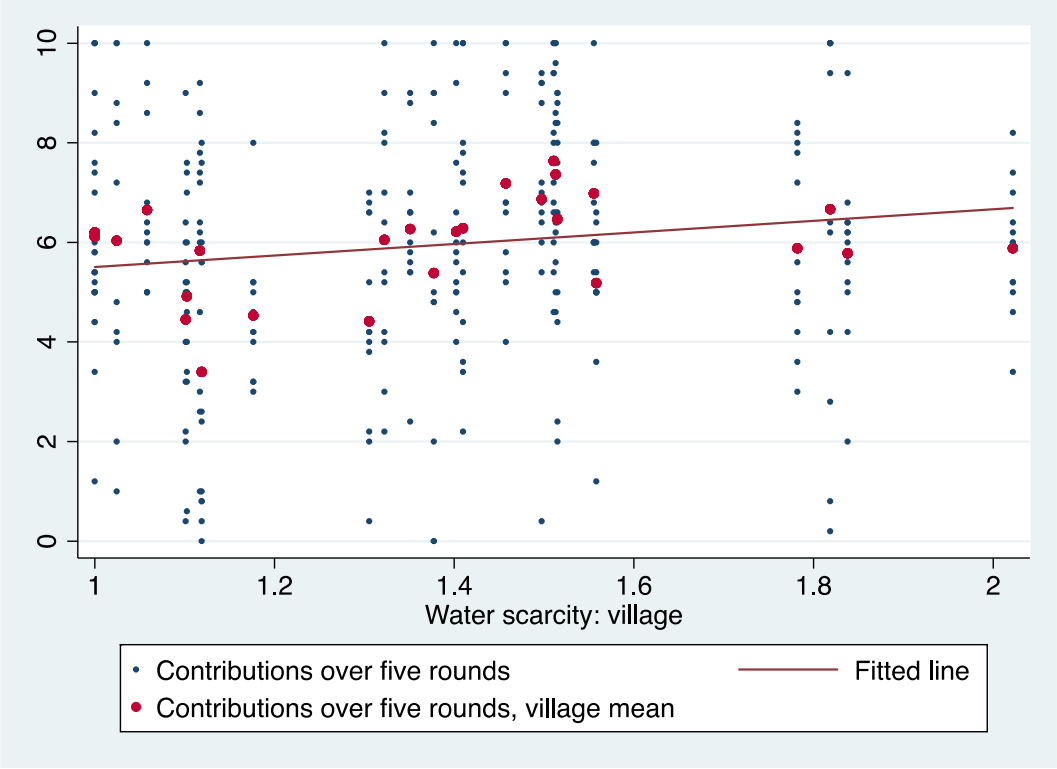


Figure 1A. The relationship between the water scarcity indicator and the average contributions in the public goods game

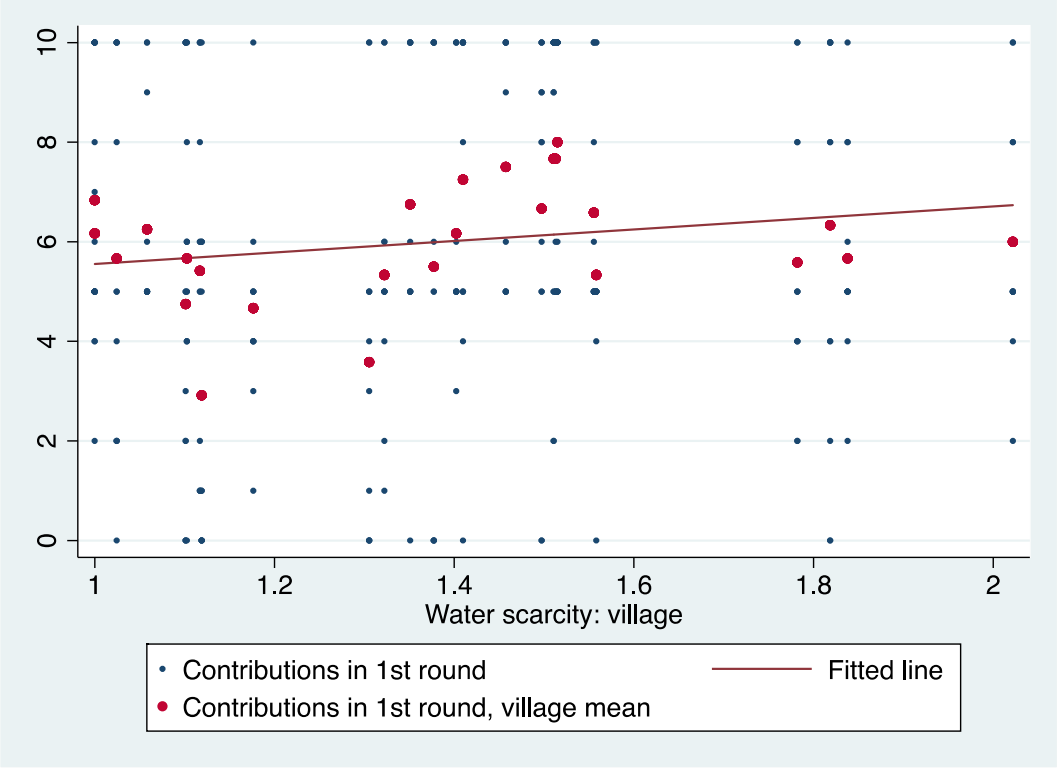


Figure 1B. The relationship between the water scarcity indicator and the contributions in the first round of the public goods game

We then formally test the relationship between water scarcity and contributions in the public goods game, using both average contributions over the five rounds and contributions in the first round as dependent variables. The OLS regression results are reported in Table 5A and 5B.[36] Regression specifications are the same in the two tables except for dependent variables. The dependent variable in Table 5A is the average contribution and the dependent variable in Table 5B is the contribution in the first round. Because the results in the two tables are similar, the following analyses are about both tables if not specified.

[36] Because contributions in the game are limited to between 0 and 10, the dependent variables are censored at both ends. We also perform tobit regressions with the same model specifications and the results are very similar. For simplicity of interpretation, we only show OLS results here; tobit regression results are available upon request.

Table 5A. Water scarcity and average contributions over five rounds of the PGG

	(1) Av. Con.	(2) Av. Con.	(3) Av. Con.	(4) Av. Con.	(5) Av. Con	(6) Av. Con	(7) Av. Con.	(8) Av. Con.	(9) Av. Con.	(10) Av. Con.	(11) Av. Con.
Water scarcity	1.159*	1.527**	1.312**	1.457**	-6.836	1.949**	0.665	1.312*	1.954**	1.045	1.043
	(0.613)	(0.658)	(0.576)	(0.608)	(5.835)	(0.911)	(0.672)	(0.647)	(0.919)	(0.698)	(0.755)
No. of wells in vill (per 1000 ha)			-0.302								
			(0.239)								
Household groundwater access			0.241								
			(0.403)								
Water scarcity, household level				0.101							
				(0.273)							
Age*water scarcity					0.163						
					(0.112)						
Not farming 2015								-2.128			
								(1.382)			
Not farming*water scarcity								1.732			
								(1.044)			
Off-farm job									1.680		
									(1.817)		
Off-farm job* water scarcity									-1.305		
									(1.255)		
Zone dummies										Yes	
Individual characteristics	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Village characteristics	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	4.346***	10.44**	11.07**	10.34**	22.35**	6.620	12.21	10.68**	10.23**	9.628**	5.805
	(0.910)	(3.894)	(4.188)	(3.875)	(8.708)	(4.978)	(8.021)	(3.872)	(3.774)	(4.030)	(5.488)
Observations	312	312	312	311	312	175	137	312	312	312	264
Adjusted R^2	0.018	0.079	0.083	0.080	0.086	0.126	0.081	0.084	0.084	0.084	0.051

Note: OLS regression results of the effect of water scarcity on average contribution over the five rounds of the PGG. The coefficients of the individual and village characteristics are omitted for the sake of brevity. Individual characteristics include age, years of schooling, size of contracted land, number of siblings, household size and dummy for having an apartment in town as a proxy for wealth. Village characteristics include natural log of village land size, dummy for having non-farm enterprises in the village, share of off-farm labor, distance to town seat and distance to county seat. Column (6) is results for a sub-sample of subjects born before 1982 and Column (7) is for subjects born in 1982 and later. Column (11) excludes the four village from Zone 3. Robust standard errors in the parentheses are adjusted for clustering at village level. ***, ** and * stand for significant level 1%, 5% and 10%, respectively.

Table 5B. Water scarcity and contributions in the first round in PGG

	(1) Fir. Con.	(2) Fir. Con.	(3) Fir. Con.	(4) Fir. Con.	(5) Fir. Con.	(6) Fir. Con.	(7) Fir. Con.	(8) Fir. Con.	(9) Fir. Con.	(10) Fir. Con.	(11) Fir. Con.
Water scarcity	1.156*	1.953**	1.738**	1.807**	-11.11*	2.567***	0.191	1.719**	2.026*	1.673*	1.961**
	(0.677)	(0.737)	(0.683)	(0.656)	(6.350)	(0.873)	(1.055)	(0.733)	(0.985)	(0.859)	(0.919)
Age		-0.0231	-0.0207	-0.0231	-0.385**	0.0616	-0.0544	-0.0264	-0.0273	-0.0228	-0.00250
		(0.0361)	(0.0355)	(0.0364)	(0.185)	(0.0916)	(0.101)	(0.0366)	(0.0352)	(0.0370)	(0.0409)
No. of wells in vill (per 1000 ha)			-0.336								
			(0.272)								
Household groundwater access			0.552								
			(0.502)								
Water scarcity, household level				0.212							
				(0.514)							
Age*water scarcity					0.254**						
					(0.120)						
Not farming 2015								-2.994			
								(1.892)			
Not farming*water scarcity								2.009			
								(1.268)			
Off-farm job									0.238		
									(2.158)		
Off-farm job* water scarcity									-0.255		
									(1.389)		
Zone dummies										Yes	
Individual characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Village characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	4.399***	13.14***	14.52***	12.92***	31.74***	7.518	15.59*	13.57***	13.33***	13.12***	9.976*
	(1.004)	(3.576)	(3.974)	(3.635)	(10.09)	(6.627)	(8.130)	(3.500)	(3.474)	(3.548)	(5.776)
Observations	312	312	312	311	312	175	137	312	312	312	264
Adjusted R^2	0.010	0.087	0.091	0.090	0.098	0.117	0.137	0.092	0.088	0.092	0.062

Note: OLS regression results of the effect of water scarcity on average contribution in the first round of the PGG. The coefficients of the individual and village characteristics are omitted for the sake of brevity. Individual characteristics include age, years of schooling, size of contracted land, number of siblings, household size and dummy for having an apartment in town as a proxy for wealth. Village characteristics include natural log of village land size, dummy for having non-farm enterprises in the village, share of off-farm labor, distance to town seat and distance to county seat. Column (6) is results for a sub-sample of subjects born before 1982 and Column (7) is for subjects born in 1982 and later. Column (11) excludes the four village from Zone 3. Robust standard errors in the parentheses are adjusted for clustering at village level. ***, ** and * stand for significant level 1%, 5% and 10%, respectively.

We first show the regression results without any controls in the first column, which confirm the relationship in Figures 1A and 1B. A higher level of water scarcity has a positive effect on cooperation among local villagers, although the significance levels are low. Column (2) shows our main results, where we add individual, household and village characteristics that may affect cooperation as additional control variables. The coefficient of water scarcity gets larger and more significant, while the coefficients of other characteristics are consistent with our expectations. Based on the size of the coefficients, one standard deviation increase in the water scarcity indicator could increase the average contributions by 0.42 Yuan, about a 7% increase from the mean, and increase the contributions in the first round by 0.54 Yuan, about a 9% increase from the mean.

As mentioned before, groundwater availability is another factor that might affect our identification strategy. Although we don't believe that the relatively recent availability of groundwater could have disturbed the long-term effects of water scarcity, we tested it in column (3). We include the number of wells per hundred hectares of arable land in the village and a dummy for access to groundwater at household level to see whether the availability of groundwater affects our result. The result confirms our previous argument. Better access to groundwater may have slightly reduced contributions, but it is not statistically significant. And the significant and positive relationship between contributions in the game and water scarcity is robust to groundwater availability.

So far, we mainly measure water scarcity at the village level, because we believe that the impact of water scarcity on cooperation works through the interaction among people and shapes people's preferences at the community level rather than at individual or household level. In column (4), we add a household-level water scarcity indicator constructed in a manner similar to the village-level indicator. The results show that household level scarcity doesn't affect contributions in the public goods game, and only village-level water scarcity is statistically significant. This finding supports our argument that individual households cannot effectively operate irrigation by themselves, and that the effect of water scarcity operates through inducing better collective irrigation management, which then fosters a more cooperative norm within the whole community.

Because we argue that a stronger norm of cooperation is formed through better collective actions, and better collective actions are induced by the higher level of water scarcity, the length of exposure to water scarcity and experience of collective actions are important for the

formation and evolution of cooperative norms. We expect that people who have been exposed to water scarcity longer and have more experience in working together as a collective should hold a higher norm in terms of cooperation. More importantly, before the land reform in 1981, villages in Minle were functioning as collectives to coordinate agricultural and irrigation activities and farmers had more experience in working together on public projects. Thus, we test the impact of exposure in two ways. First, we interact age with water scarcity in the model specification in column (5). We expect to find a positive coefficient for the interaction term. Second, we separate the sample into two subsamples based on their exposure to the collective era. We define the "old" generation as people who were at least 16 years' old in 1981 when the land reform happened, and the rest are classified as the "young" generation. We expect to find stronger cooperative norms among the "old" generation than the "young" generation. Columns (6) and (7) in Table 5 are regression results for the "old" and "young" generations respectively. The results are consistent with our expectations. The effect of water scarcity on cooperation is stronger both in the size and significance among older subjects than among the younger subjects.

Furthermore, we test the robustness of our results to subjects' current involvement in agriculture and occupational choices. In column (8), we test whether a household's engagement in agricultural production affects the effect of water scarcity on cooperation, by including a dummy variable for no farming in 2015 and its interaction term with water scarcity in the regression. In column (9), we test whether having off-farm jobs affects the impact of water scarcity on cooperation, by including a dummy variable for off-farm jobs in 2015 and its interaction term with water scarcity. The results are robust and consistent. Greater water scarcity fosters stronger norms of cooperation in rural communities and the effect is persistent and not easily disturbed by short-term events.

Finally, although our study site is limited to a small area along the Hongshui River (longest driving distance between any two villages in our sample is about 40km), there could still be some variations in local geo-climate features such as rainfall and temperature, which could affect the demand for irrigation water and thus affect the relative water scarcity. Higher rainfall and lower temperature (lower evaporation) reduce water demand and thus mitigate irrigation water scarcity. In Minle, elevation is the key determinant of local rainfall and temperature. High elevation areas have more rainfall and lower temperature and thus lower demand for irrigation water than low elevation areas. If elevation is positively correlated with our irrigation water scarcity indicator, meaning villages with relatively less water quota actually have lower

demand, the size of the water scarcity indicator might fail to capture even the direction of true water scarcity levels across villages. However, our water scarcity indicator is negatively correlated with elevation among the sampled villages (Pearson's correlation=-0.4298, p=0.0284), suggesting that villages with great irrigation water scarcity also have higher water demand. This means that differences in the size of the scarcity indicator actually understate the true difference in water scarcity. Therefore, the direction of the effect of water scarcity is not affected, although we may have overestimated the size of the effect. Because agricultural land in Minle is typically divided into three zones based on elevation, we also control for the three zones in column (10) to see whether our results are robust.[37] Although the size and the significance level of the coefficients of water scarcity decrease, the coefficient for the first-round contribution is still significant. The decrease in both size and significance level is largely due to the correlation between the zone dummies and our small sample size. Another concern about local climatic differences is that water allocation may have been adjusted to meet the water demand. If each unit of water quota equals a different amount of water in different villages, our water scarcity indicator would fail to represent the true differences in water scarcity across villages. From the Hongshui River Irrigation Administration, we only know that the amounts of water allocated to each water quota unit are the same in the 22 villages in zone 1 and zone 2 but we do not have the same level of confidence for the four villages in zone 3. Thus, in column (11), we exclude those four villages in zone 3. Again, although weaker in both magnitude and significance level due to the smaller sample size, the results are similar, especially for the contribution in the first round.

Furthermore, when we compare the results in Tables 5A and 5B, we find that the effect on the contributions in the first round of the PGG is consistently larger than the effect on the average contributions over the five rounds. This finding supports our decision-making heuristic assumption that people tend to carry the existing norms to new situations – in this case, a lab-in-the-field experiment. The effect of a cooperative norm on contribution decisions should be strongest in the first round and then may be gradually eroded by the learning process during the game play.

[37] Zone 1 has an elevation ranging from 1600 to 2000 m; Zone 2 ranges between 2000 and 2200 m; and Zone 3 ranges between 2000 and 2200 m. In our sample, among the 26 villages, 10 are located in Zone 1, 12 in Zone 2 and 4 in Zone 3.

3.4.3 IV estimation results

Although we have discussed the exogeneity of our water scarcity measure in Section 2, we cannot completely rule out potential endogeneity. Our measure could be endogenous if the water quota and thus relative water scarcity was affected by older generations' attitudes toward public goods, and if such local culture, norm or social preferences was transmitted to the younger generation. The villages that valued public goods more and thus were more willing to contribute to the canal and reservoir construction would tend to acquire more water quota than the villages that valued public goods less. This relation holds for the younger generation as well if their attitudes toward public goods are influenced by the older generation. However, if this is the case, it implies a negative relationship between water scarcity and contributions to the public goods game, which means that the OLS results in Tables 5A and 5B should have underestimated the true effect of water scarcity on contributions and this endogeneity issue does not affect our main findings.

We deal with this endogeneity concern with an instrumental variable approach. We argue that, in the context of Hongshui River Irrigation District of Minle County, the geographic location of the villages could serve as an IV for the relative water scarcity. The underlying logic is that, because the reservoir was built upstream on the Hongshui River, which is to the south of all the villages in our sample, the cost of working on the construction site was lower for the villagers who lived closer to the reservoir. Therefore, villages located in the south would be more willing to contribute than those in the north, and thus received more water quota relative to their land size and became less water scarce later on. We use latitude of each village as a proxy for villages' proximity to the reservoir and as our IV for the 2SLS regression. Because higher latitude means farther away from the reservoir, we expect to see a positive relationship between latitude and our water scarcity indicator in the first stage regression. The 2SLS results are shown in Table 6, where explanatory variables are the same as in Tables 5A and 5B. Similarly, we employ the 2SLS by using both average contribution over the five rounds and the contribution in the first round in PGG. The second stage regression results are shown in columns (1) and (2). Because the first stage is the same, we only report it once, in column (3).

As we can see, latitude is a strong predictor for the water scarcity ratio in the first stage regression, and it does not affect the contribution in the PGG if we include it along with the water scarcity indicator in the regression. Results from the second stage regression shows that,

when taking possible endogeneity of water scarcity into consideration, the effect of water scarcity on the contribution in the PGG is even stronger. Our finding that a higher level of water scarcity improves people's preference for cooperation still holds.

Table 6. Water scarcity on contributions in PGG, 2SLS results

	(1) Second stage Av. Con.	(2) Second stage First Con.	(3) First stage Water scarcity
Water scarcity	2.833** (1.168)	4.029*** (1.276)	
Age	-0.00124 (0.0310)	-0.0283 (0.0332)	0.00321** (0.00151)
Years of schooling	0.0359 (0.0474)	0.0878* (0.0514)	0.00579* (0.00307)
Contracted land (mu)	-0.0134 (0.0151)	-0.0384** (0.0164)	0.00795*** (0.00137)
No. of siblings	-0.137** (0.0611)	-0.155** (0.0747)	-0.000444 (0.00532)
Household size	0.178 (0.120)	0.265* (0.147)	-0.00558 (0.00850)
Wealth: Have an apartment	-0.478 (0.500)	-0.165 (0.633)	-0.0558 (0.0420)
ln(village arable land)	-1.041** (0.419)	-1.392*** (0.440)	0.157** (0.0745)
Dummy for enterprise in village	0.444 (0.403)	0.629 (0.486)	-0.0540 (0.0716)
Share of non-farm labor in village	-0.449 (1.457)	-1.374 (1.612)	0.0900 (0.398)
Distance to town seat	0.0301 (0.0893)	0.0575 (0.124)	0.0107 (0.0231)
Distance to county seat	-0.0172 (0.0194)	-0.0437* (0.0251)	-0.00958** (0.00365)
latitude			1.741*** (0.392)
Constant	10.74*** (3.818)	13.62*** (3.450)	-67.24*** (15.05)
Observations	312	312	312
Adjusted R2	0.025	0.026	0.448
Test statistics			
Partial F excl. instr.			19.69
KP LM stat			5.77
KP Wald stat			21.29
Endogeneity test	1.42	2.85	

Note: 2SLS regression results for average contribution over the five rounds and the contribution in the first round. The first stage regressions for the both outcome variables are the same, and thus we only report it once in column (3). Robust standard errors in the parentheses are adjusted for clusters at village level. ***, ** and * stand for significant level 1%, 5% and 10%, respectively.

3.4.4 Village level robustness check

So far, all of the analyses use individual level data. Because our water scarcity measure is on the village level, we also aggregate all the individual/household level information to the

village level and repeat the same regressions on the village level. The results are reported in Table 7 and are similar to the results in the individual-level analyses above.

Table 7. Water scarcity and contributions in the PGG, village level

	Average contribution over five rounds				Contribution in the first round			
	(1) OLS	(2) OLS	(3) OLS	(4) 2SLS	(5) OLS	(6) OLS	(7) OLS	(8) 2SLS
Water scarcity	1.159	1.656**	1.108	2.162**	1.156	1.874**	1.664	3.646**
	(0.688)	(0.765)	(1.045)	(1.078)	(0.843)	(0.869)	(1.201)	(1.462)
Age			0.0749	0.0438			0.166	0.107
			(0.158)	(0.117)			(0.182)	(0.146)
Years of schooling			-0.135	-0.237			0.0522	-0.140
			(0.298)	(0.296)			(0.342)	(0.327)
Contracted land (mu)			-0.0000873	-0.0228			-0.0445	-0.0872***
			(0.0415)	(0.0268)			(0.0477)	(0.0303)
No. of siblings			-0.0688	-0.0496			-0.175	-0.139
			(0.380)	(0.212)			(0.437)	(0.266)
Household size			1.326**	1.257***			1.370**	1.239***
			(0.534)	(0.302)			(0.613)	(0.342)
Wealth: Have an apartment			0.679	1.525			-0.0851	1.505
			(2.806)	(1.940)			(3.224)	(2.589)
ln(village arable land)		-0.816*	-1.195**	-1.193***		-1.063*	-1.411**	-1.406***
		(0.461)	(0.502)	(0.323)		(0.524)	(0.576)	(0.322)
Dummy for enterprise in village		0.272	0.138	0.146		0.322	0.327	0.342
		(0.428)	(0.472)	(0.370)		(0.486)	(0.542)	(0.419)
Share of non-farm labor in village		-0.224	-0.763	-1.015		-0.903	-1.117	-1.591
		(1.726)	(1.798)	(1.012)		(1.960)	(2.066)	(1.270)
Distance to town seat		0.0493	0.00358	0.00318		0.0809	0.0641	0.0634
		(0.0810)	(0.0830)	(0.0688)		(0.0919)	(0.0953)	(0.110)
Distance to county seat		-0.0169	-0.000301	-0.00372		-0.0408	-0.0276	-0.0341*
		(0.0224)	(0.0230)	(0.0148)		(0.0254)	(0.0264)	(0.0190)
Constant	4.346***	10.45***	6.382	7.820	4.399***	12.55***	2.773	5.477
	(0.969)	(3.638)	(8.304)	(6.327)	(1.188)	(4.129)	(9.542)	(7.987)
Observations	26	26	26	26	26	26	26	26
Adjusted R2	0.069	0.061	0.149	0.083	0.034	0.165	0.225	0.062

Note: OLS and 2SLS regression results. Column (1) - (4) are results for average contributions over the five rounds of PGG and column (5) - (8) are for the contributions in the first round of PGG. The explanatory variables are the same in Table 5. IV used in the 2SLS regressions in column (4) and (8) is latitude of the villages as used in Table 6. First stage regression results for 2SLS results are not shown in the table for simplicity and can be provided upon request. ***, ** and * stand for significant level 1%, 5% and 10%, respectively.

3.5 Concluding remarks

As the demand for natural resource has been on a constant increasing trend due to economic development and population growth, the increasing scarcity and the potential depletion of resources have become increasingly serious challenges to people's livelihood in many parts of the world. While sustainable management of resources often requires cooperation among stakeholders, it is natural to ask how resource scarcity affects people's

willingness to cooperate, especially when resource scarcity is often related to competition and conflicts. In this study, we directly examine how resource scarcity affects collective action in resource management and the general willingness of cooperation in the context of a gravity irrigation system in China. We propose that scarcity could help foster cooperation under a proper institutional arrangement. We build a simple model with coevolution of irrigation management and social norms to illustrate how water scarcity affects the norm of cooperation in rural communities. We test this idea using both experimental and survey data from an irrigation district with a historical irrigation water quota system. This irrigation water quota system formed in the 1960s creates variations in the ratio of water quota to arable land area across villages. We use these variations as the measure for degree of water scarcity. We measure cooperation in irrigation management with survey questions and the general inclination of cooperation using a lab-in-the-field experiment.

We find that water scarcity improves individuals' commitment to collective irrigation management. More importantly, the effect of water scarcity goes beyond irrigation activities. Greater water scarcity also creates a stronger norm of cooperation, measured by contributions in a public goods game. This relationship between water scarcity and preference for cooperation holds even if we take potential endogeneity into consideration, and thus this result validates the causal relationship between water scarcity and the norm of cooperation.

Our findings seem to contradict many studies on similar topics, which have found that resource scarcity often incites conflicts and competition instead of cooperation. For example, Prediger, Vollan, and Herrmann (2014) find that pastoralists in areas with lower quality grazing land are more likely to engage in anti-social behavior in an artefactual field experiment. Our explanation for this discrepancy lies in the different nature of irrigation water compared to other well-documented resources. Unlike common pool resources such as pastoral land and fisheries, which can usually be used by individuals or a small group of users, irrigation requires monetary and labor inputs at a much larger scale, which is usually beyond the reach of individuals or small groups of users. Therefore, such requirements demand collective action from local communities, and these collective actions in irrigation activities may build a norm of cooperation among local people. This feature of irrigation has been well documented in the literature (Aoyagi, Sawada, and Shoji 2014, Bardhan 2000, Bentzen, Kaarsen, and Wingender 2017, Fujiie, Hayami, and Kikuchi 2005, von Carnap 2017). As water becomes scarcer, the value of collective action increases and farmers have incentives to work more closely with each other on the irrigation system. The stronger interdependence among the farmers then shapes a

more cooperative culture, as discussed in Carpenter and Seki (2011) and Gneezy, Leibbrandt, and List (2016). Furthermore, our findings are also consistent with the literature on self-governing common pool resource management, where successful cases are often found in irrigation systems (e.g. Ostrom and Gardner 1993, Ostrom 1990).

Our findings also underscore the importance of institutions in shaping social norms and preferences. Irrigation is not only an agricultural technology, but also a set of institutional arrangements that requires users to act in a certain way in order to benefit from it. An irrigation system could foster cooperation as it strengthens the interdependence within the community. Thus, other technological or institutional arrangements with similar features should be expected to help build cooperation in a similar way. This finding is particularly pertinent in communities with rural development or common pool resource management projects. The crucial element of the success of these projects is farmers' voluntary participation and contribution, which is affected by local norms or cultural attitudes toward cooperation. If a project could include elements that enhance the experience or the perception of interdependence, there may be a better outcome in the long term because the project could create additional benefits from the more cooperative culture that it helps to foster.

Due to the design and scale of the data, this study also has its limitations. We rely on a historical irrigation water quota system to identify variations in water scarcity. This strategy has many merits that help isolate the impact of water scarcity on cooperation, as we described in Section 3. However, these merits of our empirical strategy and our measure of water scarcity do not come without a price. By focusing on one irrigation district in one county, we forgo the ability to test the heterogeneous impacts of water scarcity on cooperation. Our findings are confined to the context of the surface water irrigation system and would have a hard time when applied to other types of irrigation systems. von Carnap (2017) has provided a detailed discussion on how different irrigation systems have different impacts on social capital formation. Similarly, while we are able to rule out confounding factors such as water distribution rules, agricultural policies and social-economic history, we also lose the ability to examine how these factors could interact with water scarcity. We should be cautious in generalizing our findings to other regions with different social, economic and institutional backgrounds. Moreover, since we only have 26 villages in our sample, we don't have enough statistical power to test village-level heterogeneity effects. Studies on a larger scale are needed to answer these questions in the future.

Furthermore, our experimental design only allows us to study the inner-group cooperation. While our findings in this study emphasize the positive effects of resource scarcity on inner-group cooperation, we do not claim the same effects of resource scarcity on inter-group cooperation. People tend to act differently when interacting with people from their own group rather than with outsiders (Chen and Li 2009) and exhibit "parochial altruism" behavior, i.e., people are kind to toward members of own group but hostile towards outsiders (e.g. Abbink et al. 2012, Bernhard, Fischbacher, and Fehr 2006). The existence of threats from out-groups can increase intra-group social capital accumulation (Jennings and Sanchez-Pages 2017). Thus, when facing greater resource scarcity, and when the allocation of resources is not clearly demarcated, it is not clear how resource scarcity affects inter-group cooperation/competition. The answer to this question is beyond the scope of this study.

Another limitation of this study is that we only study the impact of water scarcity in the context of moderate water scarcity in a semi-arid area. We have to be cautious if we want to generalize the findings to different settings. When water is very scarce, the positive relationship between water scarcity and cooperation that we find in this study may not hold. The literature has recorded a curvilinear relationship between water scarcity and collective action in irrigation system (Bardhan 2000, Araral 2009, Wang, Chen, and Araral 2016). Collective action is more difficult when water is very scarce or abundant, but easier when scarcity is moderate. This means that, if we apply the same analysis to a more water-scarce area, we might find the opposite relationship between scarcity and cooperation. Again, larger-scale research with a wider spectrum of scarcity is needed to fully reveal the relationship between water scarcity and cooperation. It is left for future studies.

Appendix

Proposition 1 Proof:

Substitute $l_i = E - g_i$ and $\Delta g_i = g_i - g^N$ into the utility function and we have the following utility maximization problem: $\max_{0 \le g_i \le E} u_i = (E - g_i)^\alpha A^\beta + r_i f(g_i - g^N)$. The first order condition is: $\frac{\partial u_i}{\partial g_i} = -\alpha(E - g_i)^{\alpha-1} A^\beta + \beta(E - g_i)^\alpha A^{\beta-1} A_1 + r_i f' = 0.$

Let $\widetilde{g_\iota}$ be the short term optimal contribution of individual farmer i and satisfies the first order condition:

$$r_i f'(\widetilde{g_\iota} - g^N) = \alpha(E - \widetilde{g_\iota})^{\alpha-1} A^\beta - \beta(E - \widetilde{g_\iota})^\alpha A^{\beta-1} A_1. \quad (1)$$

Thus, $\widetilde{g_\iota}$ is a function of social norm g^N and sensitivity to norm r_i.

Taking partial derivatives of g^N on both side of the equation (1), we have

$$r_i f'' \left(\frac{\partial \widetilde{g_\iota}}{\partial g^N} - 1\right) = (-\alpha(\alpha-1)(E - \widetilde{g_\iota})^{\alpha-2} A^\beta + 2\alpha\beta(E - \widetilde{g_\iota})^{\alpha-1} A^{\beta-1} A_1 - \beta(\beta-1)(E - \widetilde{g_\iota})^\alpha A^{\beta-2} {A_1}^2 - \beta(E - \widetilde{g_\iota})^\alpha A^{\beta-1} A_{11}) \frac{\partial \widetilde{g_\iota}}{\partial g^N}. \quad (2)$$

Since $\alpha, \beta \in (0,1)$, $A_1 > 0$, $A_{11} < 0$, $r_i > 0$ and $f'' < 0$, rearranging equation (2) gives us $0 < \frac{\partial \widetilde{g_\iota}}{\partial g^N} < 1$. Short term optimal contribution $\widetilde{g_\iota}$ is increasing with social norm in regard of public good provision g^N.

Taking partial derivatives of r_i on both side of the equation (1), we have

$$f' + r_i f'' \frac{\partial \widetilde{g_\iota}}{\partial r_i} = (-\alpha(\alpha-1)(E - \widetilde{g_\iota})^{\alpha-2} A^\beta + 2\alpha\beta(E - \widetilde{g_\iota})^{\alpha-1} A^{\beta-1} A_1 - \beta(\beta-1)(E - \widetilde{g_\iota})^\alpha A^{\beta-2} {A_1}^2 - \beta(E - \widetilde{g_\iota})^\alpha A^{\beta-1} A_{11}) \frac{\partial \widetilde{g_\iota}}{\partial r_i}. \quad (3)$$

Rearranging equation (3), we can easily have $\frac{\partial \widetilde{g_\iota}}{\partial r_i} > 0$. Short term optimal contribution $\widetilde{g_\iota}$ is increasing with individuals' sensitivity to social norm r_i.

Similarly, taking partial derivatives of *W* on both side of the equation (1), we have

$$\left(r_i f'' + \alpha(\alpha-1)(E - \widetilde{g_\iota})^{\alpha-2} A^\beta - 2\alpha\beta(E - \widetilde{g_\iota})^{\alpha-1} A^{\beta-1} A_1 + \beta(\beta-1)(E - \widetilde{g_\iota})^\alpha A^{\beta-2} {A_1}^2 + \beta(E - \widetilde{g_\iota})^\alpha A^{\beta-1} A_{11}\right) \frac{\partial \widetilde{g_\iota}}{\partial W} = \alpha\beta(E - \widetilde{g_\iota})^{\alpha-1} A^{\beta-1} A_2 - \beta(\beta-1)(E - \widetilde{g_\iota})^\alpha A^{\beta-2} A_1 A_2 - \beta(E - \widetilde{g_\iota})^\alpha A^{\beta-1} A_{12}. \quad (4)$$

Therefore, $\frac{\partial \widetilde{g_\iota}}{\partial W} < 0$. Short term optimal contribution $\widetilde{g_\iota}$ is decreasing with water resource W.

Proposition 2 Proof:

The dynamic of norm g_t^N is determined by

$$\dot{g}_{t+1}^N = g_{t+1}^N - g_t^N = \int_0^{r^{max}} \widetilde{g_{\iota,t}}\, h(r)dr - g_t^N.$$

Since $0 < \frac{\partial \widetilde{g_{\iota,t}}}{\partial g^N} < 1$, $\frac{\partial \dot{g}_{t+1}^N}{\partial g_t^N} = \int_0^{r^{max}} \frac{\partial \widetilde{g_{\iota,t}}}{\partial g^N} h(r)dr - 1 < 0$. Then long-term equilibrium at $\dot{g}_{t+1}^N = 0$ is stable.

Let $\widehat{g_t^N}$ stands for its long-term equilibrium level of social norm within a community,

$$\widehat{g^N} = \int_0^{r^{max}} \widetilde{g_\iota}(\widehat{g^N}, W)\, h(r)dr. \qquad (5)$$

Taking partial derivatives of W on both side of the equation (5), we have

$$\frac{\partial \widehat{g^N}}{\partial W} = \int_0^{r^{max}} \left(\frac{\partial \widetilde{g_\iota}}{\partial g^N}\frac{\partial \widehat{g^N}}{\partial W} + \frac{\partial \widetilde{g_\iota}}{\partial W}\right) h(r)dr = \frac{\partial \widehat{g^N}}{\partial W}\int_0^{r^{max}} \frac{\partial \widetilde{g_\iota}}{\partial g^N} h(r)dr + \int_0^{r^{max}} \frac{\partial \widetilde{g_\iota}}{\partial W} h(r)dr. \quad (6)$$

Rearranging equation (6), we have

$$\frac{\partial \widehat{g^N}}{\partial W}\left(1 - \int_0^{r^{max}} \frac{\partial \widetilde{g_\iota}}{\partial g^N} h(r)dr\right) = \int_0^{r^{max}} \frac{\partial \widetilde{g_\iota}}{\partial W} h(r)dr.$$

Since $0 < \frac{\partial \widetilde{g_\iota}}{\partial g^N} < 1$ and $\frac{\partial \widetilde{g_\iota}}{\partial W} < 0$, we can easily have $\frac{\partial \widehat{g^N}}{\partial W} < 0$.

Proposition 3 Proof:

In our linear public goods game, the payoff structure is $q_i = \alpha(E - g_i) + \beta G$, with $\alpha = 1$ and $\beta = 0.5$.

If people carry this norm to other public goods situations, such as our public goods game, then we expect farmers to maximize their utility function as follows: $\max_{0 \le g_i \le E} u_i = \alpha(E - g_i) + \beta G + r_i f(g_i - g^N)$; the first order condition is: $\frac{\partial u_i}{\partial g_i} = -\alpha + \beta + r_i f'(g_i - g^N) = 0$.

Define $\Delta g_i \equiv g_i - g^N$. As $\Delta g_i \in [-g^N, E - g^N]$ and $f'' < 0$, $f'(E - g^N) \le f'(\Delta g_i) \le f'(-g^N)$.

Therefore, for anyone with $0 \le r_i < (\alpha - \beta)/f'(-g^N) \equiv r^L$, $\frac{\partial u_i}{\partial g_i} < 0$ and thus his optimal public goods contribution is $\widetilde{g_\iota} = 0$.

For anyone with $(\alpha - \beta)/f'(E - g^N) \equiv r^H < r_i \le r^{max}$, $\frac{\partial u_i}{\partial g_i} > 0$, and thus his optimal public goods contribution is $\widetilde{g_\iota} = E$.

For $r^L \le r_i \le r^H$, $\widetilde{g_\iota} = \widetilde{\Delta g_\iota} + g^N$ and $\widetilde{\Delta g_\iota}$ meets $f'(\widetilde{\Delta g_\iota}) = (\alpha - \beta)/r_i$. Let f'^{-1} be the inverse function of f', $\widetilde{g_\iota} = \widetilde{\Delta g_\iota} + g^N = f'^{-1}((\alpha - \beta k)/r_i) + g^N$.

Therefore, individual i's optimal contribution in a public goods game is

$$\widetilde{g_i} = \begin{cases} 0 & if\ 0 \le r_i < (\alpha - \beta)/f'(-g^N) \equiv r^L \\ f'^{-1}((\alpha - \beta)/r_i) + g^N & if\ r^L \le r_i \le r^H \\ E & if\ (\alpha - \beta)/f'(E - g^N) \equiv r^H < r_i \le r^{max} \end{cases}.$$

From the definition of r^L and r^H, we can easily have $\frac{\partial r^H}{\partial g^N} < 0$ and $\frac{\partial r^L}{\partial g^N} < 0$, thus $\frac{\partial H(r^L)}{\partial g^N} < 0$ and $\frac{\partial H(r^H)}{\partial g^N} < 0$, where $H(r)$ is the cumulative distribution function of r.

We can also easily know that $\frac{\partial (f'^{-1}((\alpha-\beta)/r_i)+g^N)}{\partial g^N} > 0$, therefore $\frac{\partial \int_0^{r^{max}} \widetilde{g_i} h(r_i) dr_i}{\partial g^N} > 0$.

Since we have proved that $\frac{\partial \widehat{g^N}}{\partial W} < 0$, in villages with higher level of water scarcity (thus higher $\widehat{g_t^N}$), we expect to observe higher average contribution, less people who contribute little and more people who contribute all in a public goods game.

Chapter 4 Institutional preferences, social preferences and cooperation

evidence from a lab-in-the-field experiment in rural China

Abstract: In this study, we examine institutional preferences and contribution in public goods games using a lab-in-the-field experiment with farmers in rural China. Specifically, we examine whether people act differently when the institution they prefer is implemented or not and what factors are behind subjects' institutional preferences. We find that subjects have stronger preference for the reward institution over the punishment institution. But whether subjects' preferred institution matches the exogenously assigned institution or not does not have significant impacts on their contributions in the public goods game. Moreover, we find that subjects who prefer punishment tend to be free-riders, which cannot be fully explained by strategic concern or game history. We further find that the there is a robust relationship between the preference for the punishment institution and certain efficiency-reducing social preference profiles, which may help explain the institutional preferences.

This chapter is based on:

Zihan Nie, Xiaojun Yang, Jianyin Qiu and Qin Tu (2018). Institutional preferences, social preference and cooperation: evidence from a lab-in-the-field experiment in rural China. Working Paper.

4.1 Introduction

Cooperation is vital to the success of a community. The main reason for a low level of cooperation is the existence of free-riders. Among other factors, appropriate institutions could help to reduce free-riding and thus maintain cooperation. Punishment and reward are the two most extensively studied institutional instruments in promoting cooperation. One strand of literature examines the effectiveness of exogenously imposed punishment and reward institutions (Fehr and Gachter 2002, 2000, Sutter, Haigner, and Kocher 2010, Sefton, Shupp, and Walker 2007). Another strand focuses on the effect of endogenously formed (e.g. democratic voting by the majority rule) punishment and reward institutions on cooperation (Dal Bo, Foster, and Putterman 2010, Ertan, Page, and Putterman 2009, Kamei, Putterman, and Tyran 2014, Sutter, Haigner, and Kocher 2010, Tyran and Feld 2006). Studies have shown that when given the chance to choose, people tend to avoid institutions with punishment opportunities (Botelho et al. 2005, Ertan, Page, and Putterman 2009, Gurerk, Irlenbusch, and Rockenbach 2006, Dickinson, Masclet, and Villeval 2015, Gurerk 2013, Sutter, Haigner, and Kocher 2010). Meanwhile, formal institutions, such as laws and regulations, almost entirely resort to punishment. People may react when the institution they do not like is imposed on them, e.g. in the form of civil disobedience. This raises the question that whether people act differently when the institution they prefer is implemented or not in terms of public good provision, especially when the institution is exogenously imposed on them, like many social and economic policies in developing countries. This paper aims to examine this question which has received little attention in the literature so far.

A further question is what factors are behind people's institutional preferences. Studies examining this question has been limited, and the conclusion is far from conclusive. For example, Drouvelis and Jamison (2015) examine the link between institutional choices and individual preferences such as risk aversion, loss aversion and ambiguity aversion, but find no significant relationship between them. Similarly, Kocher and Matzat (2016) do not find significant correlations between institutional preferences and social-value orientations. In this study we explain institutional choices via a comprehensive profile of social preferences. Social preferences have been found to affect cooperation (e.g. Fehr and Gachter 2002, Fischbacher and Gachter 2010, Frey and Meier 2004), but to our best knowledge there are no studies investigating the link between social preferences and institutional preferences. Therefore, the second objective of this paper is to examine whether social preferences are related with institutional preferences.

We conducted a lab-in-the-field experiment in 26 rural communities in northwestern China. The core of the experiment was a three-stage public goods game (PGG). First, subjects played a standard linear public goods game. Then, we introduced two institutions to the subjects: one centralized automatic non-deterrent punishment institution and one centralized automatic non-deterrent reward institution, and we asked subjects to indicate which institution they preferred. After reporting institutional preferences, subjects were randomly assigned to either the punishment or the reward institution, regardless of their institutional preferences. We choose exogenous and weak punishment and reward design to better mimic the institutional environment in rural China, where local lives are often affected by policies imposed by the higher authorities without democratic participation or a fair procedure, and formal institutions can often be non-deterrent due to with weak enforcement capability and high detection costs. We also elicited subjects' social preference profiles through three simplified dictator games where people choose between an equal allocation and an unequal allocation of payoffs. The combinations of these choices largely depict subjects' social preference profiles.

We find that about a quarter of all subjects prefer the punishment institution over the reward institution. Yet, whether subjects' preferred institution matches the exogenously assigned institution or not does not have significant impacts on subjects' contributions in the public goods game. This is true regardless of subjects' institutional preferences and the exogenous institutions they were assigned to. However, subjects who prefer punishment[38] tend to contribute less than those who prefer reward, which cannot be fully explained by strategic concern or the interactions that subjects face previously. We further find that the preference punishment is related to efficiency-reducing social preference profiles and that this relationship is robust to individual, household, and village characteristics and game history.

This study is related to several strands of literature. First, this study contributes to the literature investigating the effect of punishment and reward institutions on contributions in PGGs. We find that when the institutions are formal, exogenous and weak, neither punishment nor reward improve voluntary cooperation. The weak institution often merely signals what behavior is encouraged and what is discouraged. Yet, the sense of (in)appropriate behavior alone is not enough to improve cooperation among Chinese farmers. We further examine whether subjects' contributions in PGG depend on whether the exogenously imposed institution matches their institutional choices or not. We find no evidence that the mismatch

[38] In the rest of the paper, we use the terms "prefer punishment/reward" and "prefer the punishment/reward institution" interchangeably, for the sake of brevity, although their meanings are slightly different.

between exogenously imposed institutions and the preferred institution affect cooperation in the PGG. Furthermore, we use Chinese farmers as research subjects in a lab-in-the-field experiment. Compared with students and the urban population, rural redsidents in China tend to be more directly invovled in public goods provision, as traditionally they have to provide by themselves many important local public goods, such as village schools, road maintance and irrigation facilities. Our findings imply that the effects of weak institutions could be context-specific, as they contradict some findings in Vollan et al. (2017) who find significant positive effects of formal exogenous non-deterrent punishment on cooperation among migrant workers in China.

Second, our study contributes to the literature on the determinants of institutional preferences. In contrast to Drouvelis and Jamison (2015) and Kocher and Matzat (2016), this study focuses on how subjects' social preference profiles correlate with their institutional choices. We find evidence that the preference for the punishment institution is correlated to "efficiency-reducing" or "anti-social" social preference profiles. This correlation is robust to various individual, household and community features and the game history in the standard PGG. This finding mirrors the "anti-social punishment" phenomenon observed in the peer-punishment literature (Herrmann, Thoni, and Gachter 2008, Cinyabuguma, Page, and Putterman 2006, Nikiforakis 2008). Free-riders are sometimes found to punish cooperators, an action that has no direct monetary gain for both the punishers and others. Despite using a different experimental design, we observe a similar pattern where free-riders in the PGG have stronger tastes for the punishment institution. Besides strategic concern (Nikiforakis 2008) and social norms of civic cooperation (Herrmann, Thoni, and Gachter 2008), our findings suggest that social preferences may help to explain why some people choose punishment. Subjects with "efficiency-reducing" or "anti-social" preference profiles are more likely to choose punishment over reward even if they have nothing to gain from it. In a way, this finding also reflects the "pleasure of being nasty" as documented in Abbink and Sadrieh (2009).

The rest of the paper is structured as follows. In Section 4.2, we describe the experimental design and procedure. Experimental results are presented and discussed in Section 4.3. And we conclude the paper in Section 4.4.

4.2 Experimental Design and Procedure

4.2.1 Experimental design

4.2.1.1 Public goods game and institutional preference

The core of our experiment is a three-stage public goods game: first, subjects play a standard public goods game for 5 rounds; then they are asked to choose between the punishment and reward institution; finally, they play a public goods game with exogenously imposed punishment/reward institution for another 5 rounds. The order of three stages is fixed, and subjects do not have prior information about the next stage(s).

The first stage is a repeated linear public goods game without punishment or reward. The game has 5 rounds. At the beginning of each round, subjects are randomly divided into a 3-person groups and each subject received an initial endowment of 10 yuan (CNY).[39] They are then asked to allocate an integer amount of money to a group account. The payoff function can be described as follow:

$$\pi_i = 10 - g_i + 0.5 \sum_{j=1}^{3} g_j,\ i=1,2,3,$$

where π_i is subject i's payoff and g_i is his contribution to the group account. The size of the group account equals to the sum of the contribution from the three subjects in the same group. The marginal payoff of the group account is 0.5, offering monetary incentives to free ride. After all subjects make their contribution decisions, they are informed of the total group contribution and their individual payoffs before entering the next round.

After five rounds of public goods game, we introduced the formal non-deterrent punishment and reward institutions to the subjects. In the punishment (reward) institution, the payoff of the subject contributing the least (the most, respectively) in their group in a round will be automatically reduced (increased) by 1 yuan. If there are more than one subject contributing the least (the most), the 1-*yuan* punishment (reward) is split equally among them. The subjects are asked to choose an institution under which if they were to play the public goods game again.

Note that the size of punishment or reward is set to a low level so that free riding is still a Bayesian Nash equilibrium. To see this, consider first the punishment institution. Suppose everyone contributes $0 \leq k < 10$ and thus receives a punishment of 0.33 yuan. If a subject increases his contribution by 1 yuan, this decision avoids the punishment by 0.33 yuan, but reduces his payoff by 0.5 yuan. Because $0.33 - 0.50 < 0$, the subject will be better off without increasing the contribution. This argument indicates that subjects do not have incentive to

[39] At the time of the experiment,1 USD=6.55 CNY.

increase their contributions. When $k \geq 2$, one is strictly better off by deviating to a contribution of zero. Together, there are two free-riding equilibria: no one contributes and everyone contributes 1 yuan. Similar arguments apply to the reward institution and there are two free riding equilibria: one or two subjects contribute 1 yuan and the rest contribute zero. We use a weak punishment and reward design because we are particularly interested in the behavior under the presence of free-riding incentives. It is also more realistic simulation of the institutional environment in rural China, where many formal institutions suffer from weak enforcement capability and high detection costs, so that they are often non-deterrent de facto.

The punishment institution and reward institution we use in our experiment are different from the commonly used formal punishment design (see e.g. Putterman, Tyran, and Kamei 2011, Tyran and Feld 2006, Vollan et al. 2017) where there is a specific punish-free contribution threshold. We don't specify a threshold, because the existence of the threshold may invoke experimenter demand effect or make some of the subjects anchoring their contributions to the threshold. We believe our reward/punishment design also mimics many real-life situations where the appropriate or acceptable level of contribution to the public goods is not salient or the same to everyone. Our punishment/reward design is similar to the "hired gun" punishment design in Andreoni and Gee (2012), in the way that punishment (reward) targets at the lowest (highest) contributors. A difference is that we allow subjects to be punished or rewarded even if they all contribute the same amount so that punishment is strictly efficiency-reducing and reward is strictly efficiency-enhancing at group level. This automatic non-deterrent punishment and reward design also helps to avoid the strategic use of punishment/reward (Choi and Ahn 2013) or anti-social punishment (Herrmann, Thoni, and Gachter 2008) often seen in informal (peer) punishment settings.

After all subjects choose their preferred institutions, half of the groups are randomly assigned to the punishment institution and the other half to the reward institution, regardless of their stated institutional choices. Subjects are not informed of the institutional choices of others.[40] The public goods game with punishment or reward has five rounds as well. The payoff function is the same as in the standard repeated linear public goods game in the first stage except for the imposed punishment and reward as described above. Subjects are randomly

[40] We did not choose an incentivized method to ask subjects for their institutional choices for two reasons. First, most incentivized methods are complicated. Given the low education of our subjects, we believe any complication might decrease data quality. Second, because subjects did not know the tasks of later stages, subjects might state their true institutional preferences nevertheless.

divided into three-person groups at the beginning of this stage and then the group composition stayed constant. The choice of partner design is mainly due to two considerations. The first consideration is that we do not want subjects to switch between institutions, which may cause confusion among subjects. The second consideration is a practical one. We have only six subjects in each institution. Even if we use a stranger design to randomly divide six people into two groups for five rounds, the group members will hardly be strangers anyway.

After making their contribution decisions in each round, besides the total contribution of the group and the individual payoffs from the public goods game, subjects are informed of whether they have been punished or rewarded in the previous round and the size of the punishment or reward.

4.2.1.2 Social Preferences

We elicit subjects' social preferences via three binary-choice dictator games. In each dictator game choice, subjects play as dictators, and choose between an equal distribution and an unequal distribution between themselves and another randomly paired anonymous partner in the same session. Figure 1 shows the different options of the three choices.

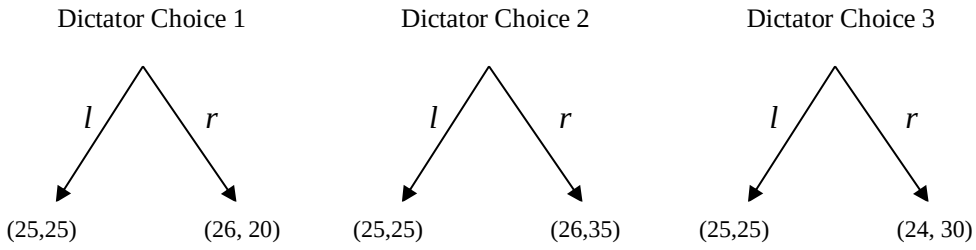


Figure 1. Binary choice dictator games for the elicitation of social preferences.

Subjects' choices boil down to the relationship among the four payoff allocations: (25, 25), (26, 20), (26, 35) and (24, 30). Different combinations of the three choices reflect different social preference profiles. We classify subjects into five mutually exclusive social preference profiles based on their choice combinations and name them accordingly to possible interpretations: subjects who always choose allocations that maximize their own payoffs are classified as "selfish"; subjects who always choose the equal allocation are defined as "inequality averse"; subjects who prefer (26, 20) over (26, 35) or prefer (24, 30) over (26, 35)

are classified as "envious"; subjects who maximize the payoffs of other in the first two choices are classified as "altruistic" [41]; and subjects who always choose the unequal allocations are defined as "equality averse".

For payment, the computer randomly pairs subjects in each dictator game and randomly selects one of the three dictator games as the basis of payment. As all subjects make decisions as dictators, one from each pair in the selected dictator choice was randomly picked as the dictator and the other as the passive receiver. The dictator's choice in the selected dictator game determined the payoffs of the both players.

4.2.2 Experimental procedure

The experiment was computerized and programed with z-Tree (Fischbacher 2007). We created a lab environment with tablets and carton boxes in the village conference rooms. We turned carton boxes into small cubicles with tablets inside so that subjects could make their decisions independently and anonymously. Communications among subjects were not permitted. Subjects received oral instructions from the experimenters at the beginning of each task and were asked to answer practice questions on paper before they started to make decisions in the experiments. The practice questions aimed to test whether the subjects understood how to calculate the payoff from the contributions in public good game. The experiment only proceeded when all subjects were able to correctly answer the practice questions. The whole experiment lasted around one and half hour. Besides the experiments, subjects also needed to finish a household survey. Subjects only got paid after they finished both the experiment and household survey. The average payment to subjects was 166 yuan, which was equivalent to one and a half days' salary of an off-farm worker.

4.2.3 Subject pool

We conducted our experiments and household survey in January 2016 in Hongshui River Irrigation District of Minle County, Gansu Province, in northwestern China.[42] We randomly

[41] This profile could be further divided into two sub-profile based on the third choice between (25, 25) and (24, 30). Subjects who are willing to give up own payoff to increase the payoff of others can be seen as "strongly altruistic" and those who are not can be regarded as "weakly altruistic". But this further classification does affect the main results.

[42] The choice of the study site is partly driven by our research interest in understanding the relationship between water scarcity and cooperation. See Chapter 2 of this thesis or Nie and Yang (2018).

selected 26 villages. In each village, we randomly 12 men as experimental subjects for one experimental session. In total, we had 312 male subjects. We deliberately chose all male subjects because men are usually household decision makers and represent their families in most of the community events.[43] Moreover, most of middle age women in rural Gansu received little education and they had trouble in understanding the setup of our experiment from experience of our pilot.[44]

The summary statistics of individual and household characteristics are described in Table 1. Ninety-seven percent of our subjects are household heads.[45] The average age is 51 years old, and the average education level is less than 7 years. About 10 percent of our subjects held positions in village administration. The subjects on average spent more than 8 months at home and they mainly engaged in the farming activities in 2015. The average net household income per capita is 14,102 yuan.

[43] It has been found that women on average have lower decision power than their husbands in the context of rural western China (Bulte, Tu, and List 2015, Yang and Carlsson 2016).

[44] Thirty-nine percent of the wives of our male subjects never received any formal school education and only seven percent received education higher than elementary school, while the corresponding figures for our male subjects are 3 percent and 57 percent respectively.

[45] One subject is the son of household head and seven are the father of the household heads.

Table 1. Summary statistics of individual and household characteristics

Variable name	Mean	SD	Min	Max	N
Individual characteristics					
Household head (dummy)	0.97	0.16	0	1	312
Age	51.4	4.96	35	66	312
Years of schooling	6.63	2.62	0	15	312
Ethnic dummy (han=1)	1	0	1	1	312
Hukou dummy (rural=1)	0.99	0.08	0	1	312
Off-farm job dummy (have any=1)	0.39	0.49	0	1	312
Time at home in 2015 (month)	8.24	3.45	1	12	312
Household characteristics					
No. of siblings	4.07	1.97	0	11	312
Majority Family name dummy (yes=1)	0.83	0.38	0	1	312
Household size (person)	4.09	1.35	1	8	312
Farm size (mu)	18.36	10.94	2.5	72	312
Gross income per capita (1000 yuan)	22.06	56.23	0.06	753.41	312
Village characteristics					
Village arable land size (mu)	4959	2265	1500	11200	312
Distance to town seat (km)	5.17	2.58	1	11	312
Distance to county seat (km)	13.79	99.51	0	40	312
Village enterprise dummy	0.35	0.48	0	1	312
Share of local nonfarm labor (%)	19.01	11.54	0.75	50.61	312
Village water quota (mu)	3807	1708	1064	9520	312
Water scarcity (ratio)	1.38	0.28	1	2.02	312
Social preferences indicators in survey					
General trust	7.28	2.07	1	10	312
Trust toward villagers	7.15	2.06	1	10	312
Take advantage of others	2.77	1.04	1	5	312
Ready to help	3.15	1.04	1	5	312

Note: Figures in the table are calculated by the authors. The *general trust* and *trust toward villagers* indicators are subjects attitude toward the statements "most people are trustworthy" and "people in the same village are trustworthy". The answers are on a 1-10 scale, 1 as absolutely disagree and 10 as absolutely agree. The *take advantage of others* and *ready to help* variables are the answers to "do you think how many people in the society will take advantage of others if given the chance?" and do you think how many people in the society are ready to help?". The answers are on a 1-5 scale, 1 as almost no one and 5 as the vast majority.

4.3 Results

4.3.1 Contributions in the public goods games

We first briefly look at subjects' average contributions and the effects of the exogenously imposed institution on contributions without considering subjects' institutional preferences. On average, subjects contributed about 60 percent of their endowment to the public account at the beginning and their contributions only slightly decreased over time. The mean contribution in the standard PGG is 5.95 yuan, close to Vollan et al. (2017) who conducted a similar PGG using Chinese migrant workers as subjects. Figure 2 shows the average group contributions in each round of the PGG. Round 6-10 in Figure 2 show average contributions of subjects in

PGGs with punishment/reward separately. For the convenience of within-subject comparison and balance check, we also show the contributions of these two groups in the standard PGG in Round 1-5 in Figure 2. We can see that contributions are similar regardless the stages of PGG and institutions implemented. We do not find any significant differences between the contributions in the PGG with punishment and the PGG with reward (clustered t-test, t=0.18, p=0.855; Somers' D, z=0.17, p=0.861). The two groups of subjects behave similarly in the standard PGG as well (clustered t-test, t=-1.42, p=0.167; Somers' D, z=-1.39, p=0.165), especially in the first round (M-W test, z=0.473, p=0.636, t-test=0.412, p=0.681). Furthermore, comparing the subjects' contributions in the PGGs with punishment/reward with their own contributions in the standard PGG, we find neither the reward nor the punishment institution has any significant effect on contributions and there is also no significant difference in the relative effectiveness of the reward and punishment institution (measured by comparing the within-subject changes in contributions of those who were assigned to either game institution).[46] Overall, the exogenously imposed non-deterrent institutions are ineffective in promoting cooperation in the PGGs.

[46] See Table A1 in Appendix for the detailed comparisons. Because we use stranger design in the standard PGG and rematch at the beginning of the PGGs with the punishment/reward institution, contributions in each session are likely related to each other. Therefore, all statistics and p-values shown in this sub-section are clustered at session level.

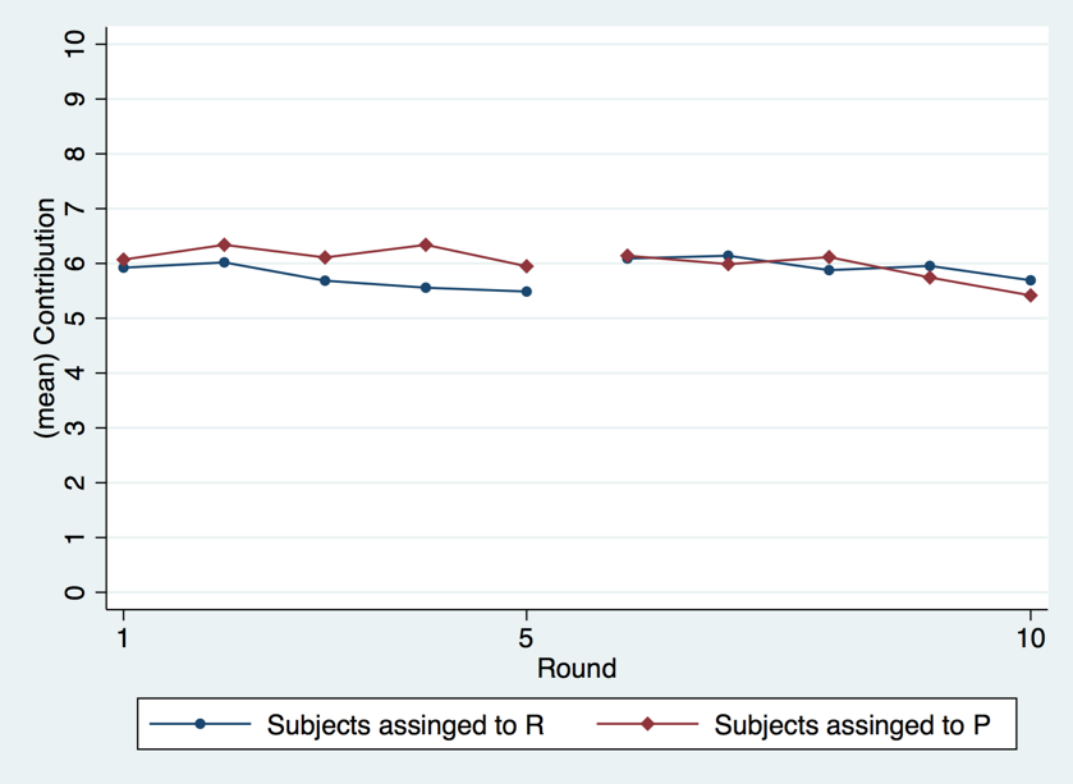


Figure 2: Evolution of contributions, by PGG institution assignment
The blue line with circle symbols is contributions of subjects who were assigned to punishment institution; the red line with diamonds is contributions of subjects were assigned to reward institution.

4.3.2 Institutional preferences, institution mismatch and contributions in the PGG

In this section, we investigate how subjects contribute under exogenously weak institutions when taking into account their institutional preferences. In particular, we aim to examine whether subjects' contribution depend on whether their preferred institutions match the exogenously assigned institutions or not.

The majority of subjects (232 out of 312) prefer the reward institution over the punishment institution. This result is also consistent with the literature on institutional preference, where people tend to avoid punishment (e.g. Drouvelis and Jamison 2015, Gurerk 2013, Gurerk, Irlenbusch, and Rockenbach 2006, Kocher and Matzat 2016, Sutter, Haigner, and Kocher 2010). Yet, this also means that about a quarter of all subjects (80/312) prefer the punishment institutions.

Because subjects are randomly assigned to groups with either the punishment or the reward institution regardless of their institutional preferences, based on which institution they prefer and which institution they are assigned to, subjects can be divided into four mutually exclusive

groups: subjects who the punishment institution and are assigned to it (45 subjects, denoted as P in P); subjects who prefer punishment but are assigned to the reward institution (35 subjects, denoted as P in R); subjects who prefer reward but are assigned to the punishment institution (111 subjects, denoted as R in P); and subjects who prefer the reward institution and are assigned to it (121 subjects, denoted as R in R).

Figure 3 shows the development of the contributions of four groups subjects in both the standard PGG and the PGGs with exogenous institutions, where Round 1-5 are the contributions of the four aforementioned groups in the standard PGG and Round 6-10 are the contributions of the four groups in PGG with exogenous institutions.

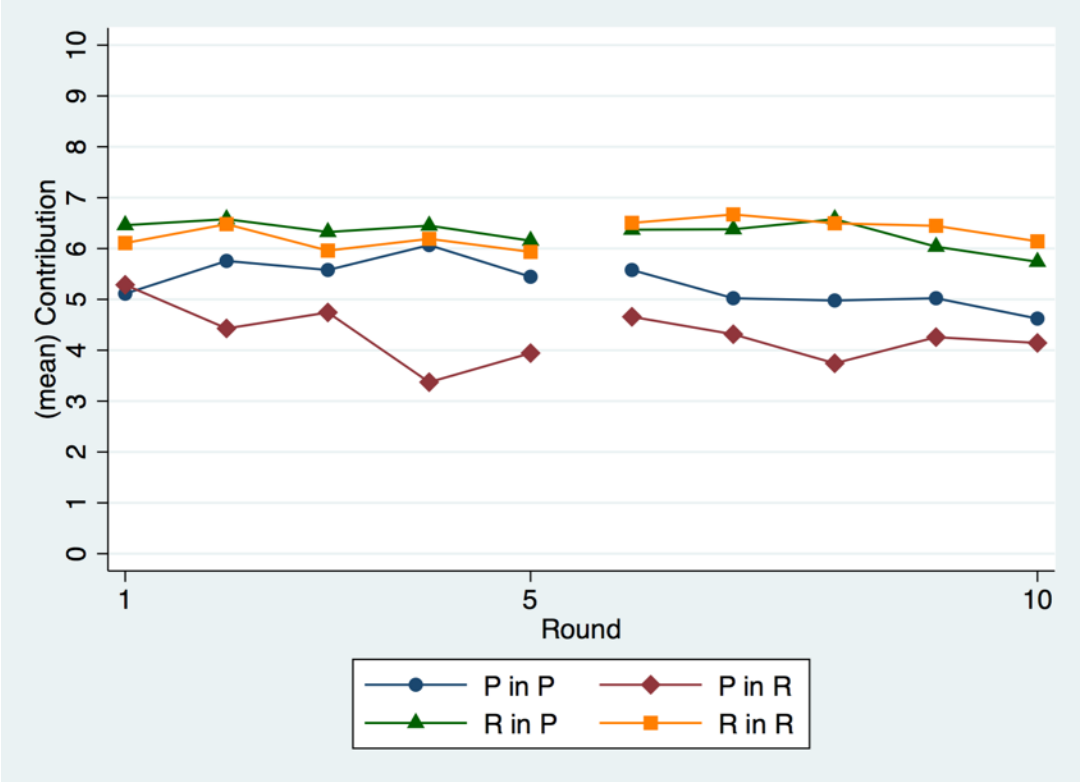


Figure 3. Mean contribution over 5 rounds of different stages of PGGs, by the combinations of institutional preferences and institution assignment.
The blue line with circle symbols is contributions of subjects who chose punishment and were actually assigned to punishment institution (P in P); The red line with diamonds is contributions of subjects who chose punishment and were actually assigned to reward institution (P in R); The green line with triangles is contributions of subjects who chose reward and were actually assigned to punishment institution (R in P); The orange line with squares is contributions of subjects who chose reward and were actually assigned to reward institution (R in R).

First, we compare the contributions of subjects who are assigned to their preferred institution with those who are not (Round 6-10 in Figure 3). For subjects who prefer the reward

institution, those who are assigned to punishment (R in P) contribute similarly to those are assigned to reward institution (R in R). For subjects who prefer the punishment institution, those who are assigned to their less preferred reward institution (P in R) contributed slight less than those were assigned to their preferred punishment institution (P in P), although the difference is not statistically significant (clustered t-test, t= -1.38, p=0.182, Somers'D, z= -1.46, p=0.146).

Yet notice that, in the standard PGG (round 1-5 in Figure 3), although subjects are randomly assigned to either punishment or reward institution and there is little prior difference in the first round (P in P vs. P in R, cluster t-test, t=-0.17, p=0.868; R in P vs. R in R, cluster t-test, t=0.69, p=0.494), there are sizable differences in contributions between subjects in "P in P" and "P in R" in later rounds and the differences can be ignored (overall contribution gap = 1.24, clustered t-test, t=2.31, p=0.009; Somers'D, z=2.37, p=0.018).[47] Therefore, between-subject comparisons are not enough to assess the impact of a mismatch between institutional preferences and exogenously assigned institutions. We adopted a difference-in-difference analysis shown in Table 2. First, we make within-subject comparisons to see how subjects change their contributions from the standard PGG to the PGG with punishment or reward for each of the four aforementioned groups. Second, we perform a between-subject comparison to examine differences in contribution changes between subjects who are assigned to their preferred institution and those who are not.

Table 2 shows the results using both the contributions in the first round and the average contributions over the five rounds as outcome variables. In column (4), we examine whether there are institutional mismatch effects for subjects who prefer the same institution but are exogenously assigned to different institutions. In the first round, we find some suggestive evidence that institutional mismatch may play a role. Subjects who are assigned to their preferred institutions contributed more than those who were assigned otherwise, for both subjects who prefer punishment and those who prefer reward. And there are sizable negative mean differences for institutional mismatch (especially for subjects who prefer punishment). Yet, none of these differences are statistically significant. Furthermore, if we look at the average contribution of the five rounds, the institutional mismatch effect is similar for people who prefer reward, but the direction is reversed for those who prefer punishment. Again, the

[47] We also test the differences in round 1-5 round by round. The gaps between P in P and P in R are significant in round 2, round 4 and round 5, especially in round 4. There are no statistical differences in contributions between R in R and R in P is any round.

differences are not statistically significant. In column (5), we also look at the institution mismatch from a different angle, examining whether there are differences in contribution for subjects with different institutional preferences but under the same exogenous institution. The results are similar. There are relatively sizable differences consistent with negative effects of institution mismatch in the first round, but they are not statistically significant. These results suggest that although subjects favor one institution over the other, they do not respond much when the institutions are exogenous imposed on them.

Table 2. Contributions conditional on institutional choices and exogenously imposed institutions

	(1)	(2)	(3)	(4)	(5)
	Standard PGG	PGG with P/R	(2)-(1) Mean diff. (t-statistics)	Institution mismatch effect (t-statistics): Same preference, different assignment	Institution mismatch effect (t-statistics): Different preference, same assignment
Contributions in the first round					
Subjects prefer P, assigned to P (P in P)	5.111	5.578	0.467 (0.37)	Mean diff. P in R – Mean diff. P in P: -1.095 (-0.71)	Mean diff. R in P – Mean diff. P in P: -0.557 (-0.37)
Subjects prefer P, assigned to R (P in R)	5.286	4.657	-0.629 (-1.01)		
Subjects prefer R, assigned to P (R in P)	6.459	6.369	-0.090 (-0.14)	Mean diff. R in P – Mean diff. R in R: -0.487 (-0.63)	Mean diff. P in R – Mean diff. R in R: -1.025 (-1.33)
Subjects prefer R, assigned to R (R in R)	6.107	6.504	0.397 (0.90)		
Contributions over five rounds					
Subjects prefer P, assigned to P (P in P)	5.591	5.044	-0.547 (-1.30)	Mean diff. P in R – Mean diff. P in P: 0.415 (0.89)	Mean diff. R in P – Mean diff. P in P: 0.374 (0.82)
Subjects prefer P, assigned to R (P in R)	4.354	4.223	-0.131 (-0.50)		
Subjects prefer R, assigned to P (R in P)	6.393	6.22	-0.173 (-0.54)	Mean diff. R in P – Mean diff. R in R: -0.490 (-1.27)	Mean diff. P in R – Mean diff. R in R: -0.449 (-1.31)
Subjects prefer R, assigned to R (R in R)	6.134	6.451	0.317 (1.46)		

Note: Figures in this table are calculated by the authors using difference-in-difference style regressions, controlling individual fixed effects and round fixed effect. Standard errors clustered at village level to calculate t-statistics. *** if p<0.01, ** if p<0.05 and * if p<0.10.

Then, the question is why subjects have preference for certain institution, especially the preferences for the inferior punishment institution?

Figure 3 reveals some interesting results when we compare the contributions of subjects who prefer the reward institution (R in P and R in R) with the contributions of those who prefer the punishment institution (P in P and P in R). Subjects who prefer punishment contribute consistently less than those who preferred reward in both the standard PGG (5.05 yuan vs. 6.26 yuan, clustered t-test, t=-3.95, p=0.001; Somers' D, z=-3.80, p=0.000) and the PGGs with an exogenous institution (4.69 yuan vs. 6.34 yuan, clustered t-test, t=-4.37, p=0.000; Somers' D, z=-4.53, p=0.000). As the punishment institution is designed to automatically punish the lowest contributors, it is surprising to see that a preference for the punishment institution is actually associated with lower levels of contributions.

This result is inconsistent with findings in the literature on PGGs with peer punishment/reward where subjects strategically choose the punishment institution to improve contributions. First, our formal automatic non-deterrent punishment/reward institutions have little space for such strategic considerations. Second, if subjects choose the punishment institution to improve cooperation, they should increase their own contributions when assigned to the PGG with punishment. But, as our results suggest, subjects do not respond to the punishment institutions significantly, even for those who prefer punishment.

Because we elicit institutional preferences after the five-rounds of standard PGG, interaction in the standard PGG may affect which institution subjects prefer and provide an alternative explanation for the negative relationship between contributions and the preferences for punishment. Subjects who are matched with uncooperative group members and fall victims of free-riding may want to punish free-riders and lower their contributions at the same time. Thus, both the lower average contributions from punishment-lovers and the preference for punishment could be the results of the misfortune in the initial interactions.

This explanation is not supported in our data. Figure 4A and 4B show subjects' own contributions and payoffs as well as the average contributions and payoffs of other group

members in each round of the standard PGG by subjects' institutional preferences.[48] As we can see, the lower contributions of subjects who prefer punishment are not driven by the misfortune of encountering free-riders. On the contrary, we see that the average contributions of subjects who prefer punishment are lower not only in absolute terms, but also in relative terms. Subjects who prefer punishment consistently contribute less and have higher payoffs than other members in their groups. And in the first round, the average contributions of other group members are similar for both subjects who prefer punishment and those who prefer reward and the main difference comes from subjects' own contributions. It means that subjects who prefer punishment didn't match with free-riders, but they were the free-riders who pulled down the group contributions.

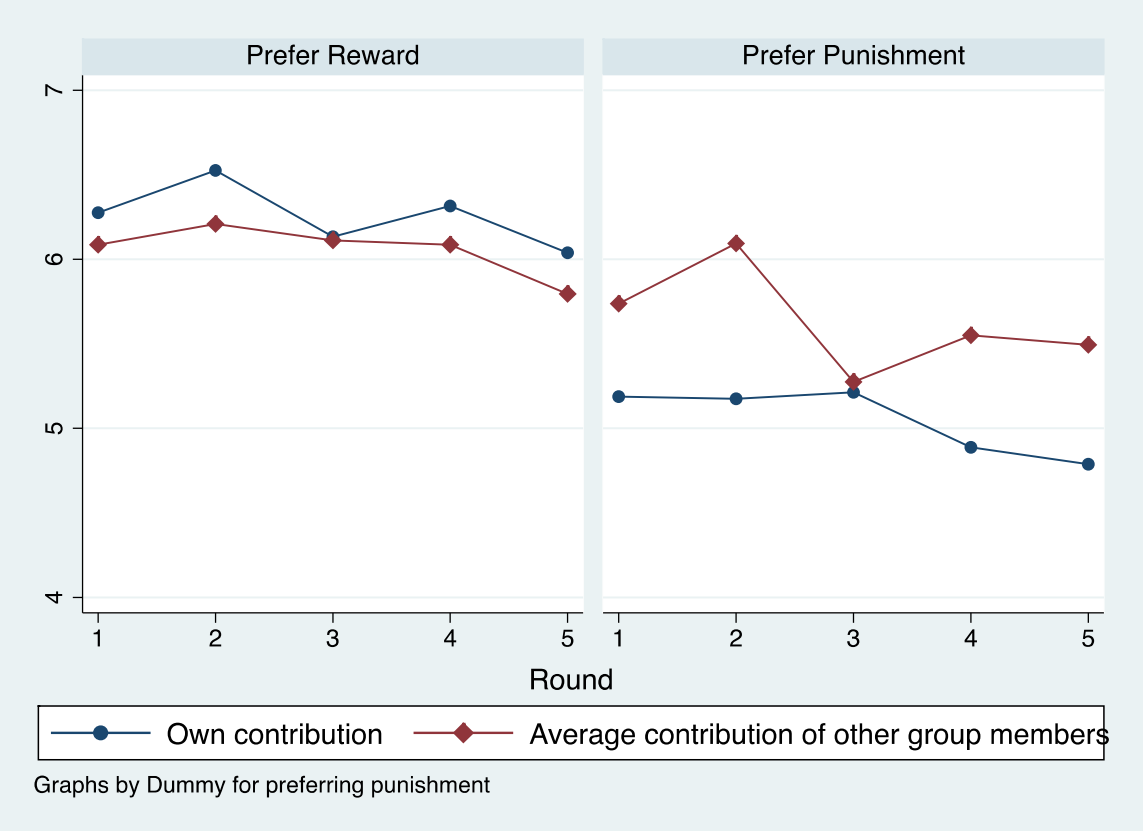


Figure 4A. Contributions in each round of the standard PGG, by institutional preferences
The figure panel on the left are for the subjects who preferred the reward institution and the figure panel on the left are for the subjects who preferred the reward institution.

[48] Detailed information about the contributions of subjects with different institutional preferences in the PGGs are shown in Table A2 in the Appendix.

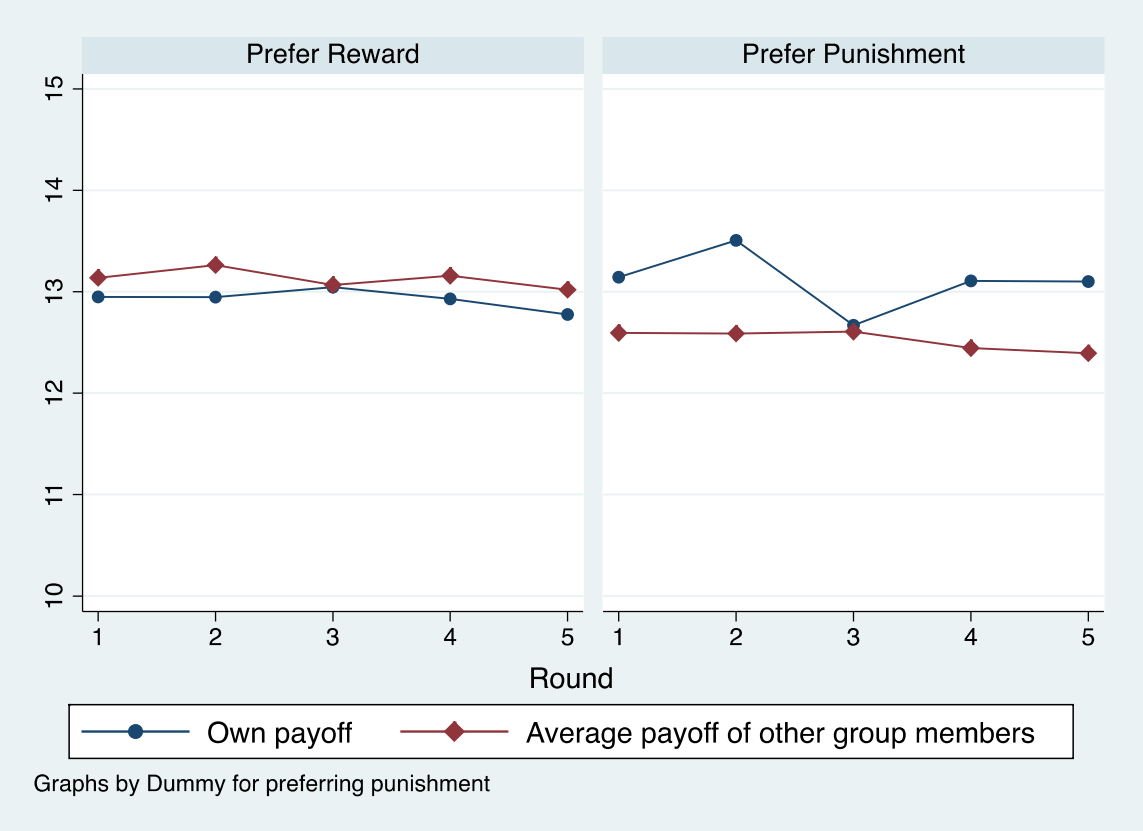


Figure 4B. Payoffs in each round of the standard PGG, by institutional preferences
The figure panel on the left are for the subjects who preferred the reward institution and the figure panel on the left are for the subjects who preferred the reward institution.

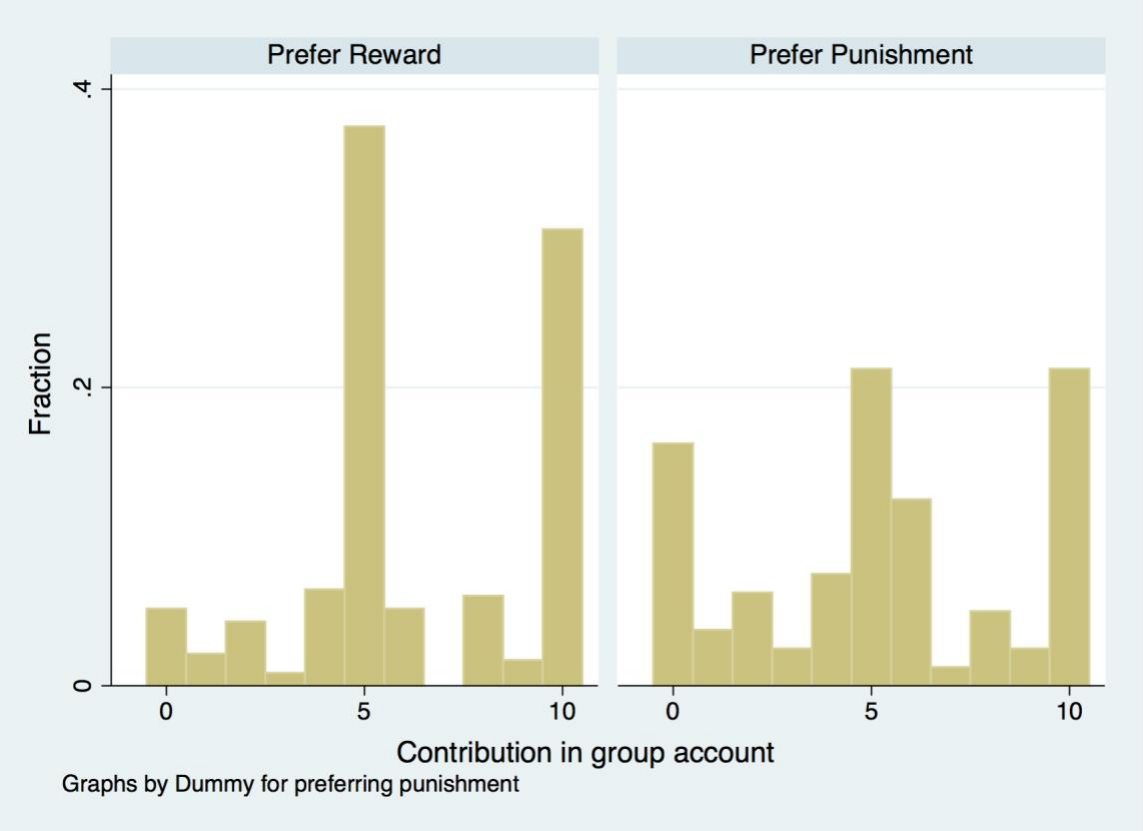


Figure 5. Distribution of first-round contributions in standard PGG
The figure panel on the left are for the subjects who preferred the reward institution and the figure panel on the left are for the subjects who preferred the reward institution.

Contributions in the first round of the standard PGG are not affected by the game history and can be considered as a measure of people's willingness of cooperation. In Figure 5, we further look at the distribution of first-round contributions conditional their institutional preferences. One result stands out clearly: there is a much higher percentage of free-riders (those who contribute nothing) among subjects who prefer the punishment institution. Based on the contribution distribution in Figure 5, we classify the whole sample into three categories according to their contribution level in the first round of standard PG: "High" contributors are subjects who contribute at least 7 yuan; "Middle" contributors are those who contribute 4-6 yuan; and "Low" contributors are those who contribute no more than 3 yuan. We show the proportions of people who prefer the punishment institution in each of the three categories in Table 3. We find that there is more than 44% of "Low" contributors prefer the punishment institution, much higher than the proportion among the "High" and "Middle" contributors. The above findings imply that people who prefer punishment are less cooperative to begin with and the game history is not enough to explain the seemingly unreasonable preferences for the

punishment institution. It is thus intriguing to know what else are behind people's preference for the punishment institution.

Table 3. Contributions in the first round of the standard PGG and institutional preferences

	(1)	(2)	(3)
Contribution in the first round in stand PGG	No. of Subjects	No. of subjects who prefer the punishment institution	% in each type
Low	52	23	44.2%
Medium	147	33	22.4%
High	113	24	21.2%
	312	80	25.6%

Note: We define subjects who contributed from 0 to 3 *yuan* in the first round of the standard PGG as low contributors (Low); those who contributed from 4 to 6 *yuan* as medium contributors (Medium); and those who contributed from 7 to 10 *yuan* as high contributor (High).

4.3.3 Social preferences and preference for punishment

In this section, we investigate the relationship between subjects' social preferences and institutional preferences. As described in Section 2, we elicit subjects' social preferences with three binary choice dictator games. Based on subjects' choices in those three games, we classify subjects into five mutually exclusive social preference profiles and name them accordingly to possible interpretations. The distribution of these five social preference profiles in our sample is reported in Table 4. The "altruistic" profile and the "inequality averse" profile are the two most common social preference profiles, taking up about 65% of the sample, while the "selfish" profile has the smallest number of subjects in all the five profiles.

Different preference profiles have different implications in terms of social efficiency. The "selfish" and "inequality averse" profiles don't care about the social efficiency. The "envious" profile seems to be efficiency-reducing or anti-social, as people are willing to reduce others' payoffs without personal gains or even willing to sacrifice own payoffs. Meanwhile, the "altruistic" profile are more pro-social or other-regarding, as people refrain themselves from take advantage of others and are willing to accept a large disadvantageous inequity for a mutually beneficial allocation. The "equality averse" profile is a bit tricky to interpret. But notice that by choosing (24, 30) over (25, 25), these subjects are willing to trade their own 1 *yuan* for an increase in the payoff of others by 5 *yuan*, while by choosing (26, 20) instead of (25, 25), they refuse to make the same trade. This relationship implies a "perverse" utility function, where people have an increasing rate of substitution between own payoff and others' payoff as they are more willing to trade own payoff for the benefits of others when they have

less, but less willing to do so when they are better-off. Or reversely, they are more willing to hurt others for the benefit of their own when they are richer than when they are poorer. In a way, it can also be perceived as an "anti-social" type. Because our punishment institution is efficiency-reducing and the reward institution is efficiency-enhancing given contributions in the PGG, we would like to examine whether the preferences for the punishment institution are related to the social preference profiles with "efficiency-reducing" or "anti-social" traits.

Table 4 also shows the proportion of subjects who prefer the punishment institution in each social preference profile. We observe substantial heterogeneities in the proportion of subjects who prefer the punishment institution across different social preference profiles, ranging from 8% to 41%. Subjects of "envious" and "equality averse" profiles have a much stronger preference for the punishment institution than others. About 40% of subjects of both profiles preferred the punishment institution[49], whereas people of all the others profiles preferred the punishment institution in less than 21% cases. It seems that the preferences for punishment are related to "efficiency-reducing" or "anti-social" social preferences.

Table 4. Social preference profiles from the three dictator games and institutional preferences

		(1)	(2)
Preference profiles	Choice combination	No. of Subjects	No. (%) of subjects who prefer the punishment institution
Inequality aversion (IA)	(l, l, l)	92	16 (17.4%)
Altruism (AL)	(l, r, l), (l, r, r)	110	23 (20.9%)
Envious (EV)	(l, l, r), (r, l, l), (r, l, r)	63	26 (41.3%)
Selfish (SE)	(r, r, l)	12	1 (8.3%)
Equality aversion (EA)	(r, r, r)	35	14 (40.0%)
		312	80 (25.6%)

Note: Figures are calculated by the authors.

We further run Probit regressions to examine whether these relationships are robust when taking other factors into account. The results are shown in Table 5 and the figures reported are the marginal effects. In column (1), we only include dummy variables for different preference profiles with the "altruistic" profile as reference group. We then add individual, household and

[49] As shown in Table 4, the "envious" profile contains three different choice combinations. However, the proportions of subjects who prefer punishment are very similar (41% in (l, l, r), 42% in (r, l, l) and 41% in (r, l, r)). Therefore, we do not further discuss them separately.

village characteristics in column (2).[50] In column (3), we further include survey based social preference indicators (trust toward villagers, whether people are willing to take advantage of others, and whether people are ready to help others). To further control for the possible influence of game history in the standard PGG, in column (4), we add subjects' contributions and the gap between own contribution and the average contribution of other group members in the first round of the standard PGG as additional explanatory variables. To distinguish possible different effects of being free-ridden by others and free-riding others, we include the size of positive gap and the absolute value of the negative gap respectively.

The main findings are robust to the inclusion of various additional explanatory variables and consistent with the descriptive results in Table 4. With the "altruistic" preference profile as the reference group, the two "anti-social" preference profiles, namely "envious" and "equality averse" still stand out. Subjects of the "envious" profiles are about 15 percentage points (specific size depends on which regression specification we look at) more likely prefer the punishment institution than those of "altruistic" preference profile. This result implies that the preference for the punishment institution is indeed related to certain efficiency-reducing or anti-social preferences. Interactions in the first round of the standard PGG also have explanatory power for the preference for the punishment institution. Given own contributions, the experience of being free-ridden by other others (measured by the positive gap between own contribution and average contribution of others) increases the likelihood of subjects preferring the punishment institution, while the experience of being free-riders does not have additional effect on the likelihood of choosing the punishment institution. Although including game history indicators partly offsets the relationship between "efficiency-reducing" preferences and the preference for punishment (decreasing in the size and the significance level, especially for the "equality averse" profile), it does not fully account for the preference for punishment, as we argued in section 3.2. The experience of being free-ridden can increase the likelihood of choosing punishment, it is more likely to affect people who make high contributions, not the free-riders. It helps explain why some of the "Middle" and "High" contributors in Table 3 prefer punishment, but not the "Low" contributors. Adding game history variables in the regression does not affect the relationship between the institutional preference and the social

[50] The individual and household characteristics variables include age, years of schooling, number of siblings, household land size, household size and whether the household owns an apartment in town (as proxy for wealth). The village characteristics variables include village land size, dummy for village enterprise, share of off-farm labor, distance to nearest town seat and distance to the county seat.

preferences much. The preference for punishment is still significantly correlated with the social preferences with "efficiency-reducing" or "anti-social" features.

Table 5. Social preference profiles and preference for the punishment institution

	(1)	(2)	(3)	(4)
Social preference profiles				
Inequality averse	-0.0394	-0.0221	-0.0256	-0.0492
	(0.0558)	(0.0574)	(0.0599)	(0.0588)
Selfish	-0.175	-0.232	-0.269	-0.256
	(0.185)	(0.181)	(0.168)	(0.167)
Envious	0.180***	0.175***	0.167***	0.145**
	(0.0543)	(0.0580)	(0.0581)	(0.0601)
Equality averse	0.170*	0.157*	0.140*	0.130
	(0.0875)	(0.0841)	(0.0823)	(0.0854)
General trust			0.0170	0.0188
			(0.0136)	(0.0133)
Trust toward villagers			0.00492	0.00548
			(0.0129)	(0.0129)
Take advantage of others			0.0367	0.0359
			(0.0233)	(0.0229)
Ready to help			-0.00656	-0.00684
			(0.0237)	(0.0234)
1st round Contri. in Standard PGG				-0.0245*
				(0.0144)
Positive gap in 1st round Con. (own-other)				0.0250*
				(0.0147)
Negative gap in 1st round Con. | own-other |				0.0121
				(0.0169)
Individual and household characteristics		Yes	Yes	Yes
Village characteristics		Yes	Yes	Yes
Observations	312	312	312	312
Pseudo R^2	0.051	0.095	0.108	0.136

Note: Robust standard errors clustered at village level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Marginal effects are reported. The "altruistic" preference profile is used as the reference group for social preference profile dummies. The individual and household characteristics variables include age, years of schooling, number of siblings, household land size, household size and whether the household owns an apartment in town (as proxy for wealth). The village characteristics variables include village land size, dummy for village enterprise, share of off-farm labor, distance to nearest town seat and distance to the county seat. The *general trust* and *trust toward villagers* indicators are subjects attitude toward the statements "most people are trustworthy" and "people in the same village are trustworthy". The answers are on a 1-10 scale, 1 as absolutely disagree and 10 as absolutely agree. The *take advantage of others* and *ready to help* variables are the answers to "do you think how many people in the society will take advantage of others if given the chance?" and do you think how many people in the society are ready to help?". The answers are on a 1-5 scale, 1 as almost no one and 5 as the vast majority. The risk choice variable is a dummy variable that equals 1 if subjects choose the risky option in the risk choice game.

4.4 Conclusion

There exists an extensive literature on the efficacy of punishment and reward institutions in promoting cooperation in a public goods game setting, and the tendency of avoiding punishment institutions when subjects are allowed to choose (Dickinson, Dutcher, and Rodet

2015, Drouvelis and Jamison 2015, Gurerk 2013, Kocher and Matzat 2016, Sutter, Haigner, and Kocher 2010). This study examines how the institutional preferences interact with exogenously assigned institutions and what factors affect a person's institutional preference, especially the roles of social preferences.

We conducted a lab-in-field experiment with 312 male subjects in rural northwestern China. We introduced exogenously imposed centralized punishment and reward institutions in public goods game settings and asked for subjects' preference over the two institutions before randomly assigning them to either of the institutions. In this way, some subjects' institutional preferences matched the institutions they were assigned to, while for others they did not match. Thus, we can examine whether people's behavior in the PGG depends on whether their preferred institution matched with the exogenously assigned institution or not. We adopted a formal exogenous and non-deterrent punishment/reward design that is meant to resemble real-life situations in rural China. We also elicit subjects' social preference profiles using three binary choice dictator games and examine whether social preference profiles are related with institutional preferences.

We find that the exogenously imposed institutions, both the punishment and the reward institution, do not have significant impacts on subjects' contributions in PGGs. Furthermore, we do not find any statistically significant effects of institution mismatch on contributions, neither among subjects who prefer the punishment institution nor among those who prefer the reward institution. Subjects who are assigned to the institution they do not prefer do not behave differently from those who are assigned to the institution they prefer. Finally, we find that subjects who prefer the punishment institution contribute significantly lower than those who prefer the reward institution. In particular, "low" contributors are more likely to prefer punishment. Neither strategic concerns nor game history can fully explain subjects' preference for the punishment institution, nor can they explain why subjects who prefer the punishment contribute less than those who prefer the reward institution. With subjects' social preference profiles elicited from three binary choice dictator games, we find that a preference for the punishment institution is related to certain "efficiency-reducing" or "anti-social" social preference profiles. This relationship is robust to the inclusion of various individual, household and village characteristics and to the game history indicators. This finding provides some additional insights into the "anti-social" punishment behavior identified in the literature (Herrmann, Thoni, and Gachter 2008, Thoni 2014). People may choose to punish others not only because of strategical concerns or for enforcing social norms, but also because of certain

intrinsic "anti-social" preferences or for simply enjoying the "pleasure of being nasty" (Abbink and Sadrieh 2009).

The lack of response to the exogenously imposed institutions may the results of the typical top-down governing regimes in rural China. Farmers who are used to having polices imposed on them may have learnt not to take their own preferences into consideration. In terms of policy making it implies that, in certain context, exogenously policies or policy changes may have may have a weak effect. Moreover, our findings may also provide some insights into endogenous policy-making process. Active support of a particular institution or policy from certain groups in the society does not necessarily represent their commitment to socially desirable outcomes that the institution or policy in question aims to achieve, and their willingness to comply if it is ever implemented. The preference for an institution could be driven by underlying preferences, beliefs or ideologies, instead of the beliefs in the effectiveness of such institution in achieving the goals.

While this study offers some insights into the relationship between institutions, preferences and cooperation, we would like to be cautious when generalizing the results to other situations. The punishment and reward institutions in our experiment are designed to be formal (from the experimenter, not other subjects), non-deterrent (the incentive of free-riding) and exogenously imposed (regardless their institutional preferences). The lack of response to the institution mismatch might not apply to other institutional settings. Future studies should expand to alternatives institutions and examine whether the relationship between institutional preferences and the effect of implemented institutions still hold, especially for endogenously determined institutions. Moreover, when using farmers as experimental subjects, we are looking at farmers carrying experiences from their own context. The context of irrigation agriculture may contribute to non-impact findings, because irrigation is found be to associated with concentration of power and less democratic attitudes (Bentzen, Kaarsen, and Wingender 2017). People from other background could react differently, which is left for studies in the future. Finally, the relationship between preferences for the punishment institution and "anti-social" or "efficiency-reducing" social preference profiles is only correlational, not causal. Future studies are needed to uncover the mechanisms behind this correlation.

Appendix

A1. Effectiveness of the reward and punishment institution in promoting cooperation in PGG

	(1)	(2)	(3)	(4)
	Standard PGG	PGG with P/R	(2)-(1) Mean difference (t-statistics)	Diff. in mean difference (t-statistics)
Contribution in the first round				
Subjects assigned to R	5.923	6.090	0.167 (0.47)	-0.096 (-0.15)
Subjects assigned to P	6.071	6.141	0.071 (0.13)	
Contribution over five rounds				
Subjects assigned to R	5.735	5.951	0.217 (1.21)	-0.497 (-1.51)
Subjects assigned to P	6.162	5.881	-0.281 (-1.01)	

Note: Figures in this table are calculated by the authors using difference-in-difference regressions, controlling individual fixed effects and round fixed effect. Standard errors clustered at village level are used to calculate t-statistics. *** if p<0.01, ** if p<0.05 and * if p<0.10.

A2. Subjects' contributions conditional on their institutional preferences

Round	Av. Contributions in Standard PGG		Mann-Whitney test z-statistics	Av. Contributions in PGG with an exogenously imposed institution		Mann-Whitney test z- statistics
	Choose reward (232)	Choose punishment (80)		Choose reward (232)	Choose punishment (80)	
1	6.28 (34.2%)	5.19 (30.8%)	2.246**	6.44 (35.1%)	5.17 (28.3%)	3.129***
2	6.53 (34.5%)	5.18 (30.1%)	3.089***	6.53 (34.8%)	4.71 (29.2%)	4.607***
3	6.13 (33.3%)	5.21 (33.5%)	1.885*	6.53 (35.4%)	4.44 (27.3%)	5.343***
4	6.31 (33.9%)	4.89 (31.6%)	3.163***	6.25 (34.7%)	4.69 (29.3%)	3.377***
5	6.04 (34.2%)	4.79 (30.2%)	2.935***	5.95 (34.9%)	4.41 (28.8%)	3.648***
Total	6.26 (34.1%)	5.05 (31.2%)	6.003***	6.34 (35.0%)	4.69 (28.6%)	8.926***

Note: Calculated by the authors from the experimental data. Figures in the parentheses are the average proportion of one's contribution in the total contribution of the group in given round. *** if p<0.01, ** if p<0.05 and * if p<0.10.

Chapter 5 Leading-by-example, leadership legitimacy and cooperation

evidence from sequential move public goods games

Abstract: Leading-by-example is a potential mechanism to improve cooperation in social dilemma situations. This study examines the impact of leading-by-example and the leadership legitimacy on group cooperation and on leaders' behavior under a sequential move public goods game setting using a novel experiment design. We find that having a leader improves cooperation in public goods games. The increase of group contributions is induced by the leader's contributions. Being a leader makes leaders increase their contributions, and followers reciprocate but to a slightly smaller extent and thus they harvest the gains from increased group contributions. The perception of leadership legitimacy does not have additional impacts on the leaders' contribution level and the group contribution level. But we find evidence suggesting that leadership legitimacy influences how leaders make their contributions decisions and update their beliefs in the repeated public goods game. Only "legitimate" leaders show the strategic use of "leading-by-example", not "appointed" leaders, whose higher contributions as leaders are simply a reaction to the sequential move game structure where their behavior can be observed by others.

This chapter is based on:

Zihan Nie and Niccolò Meriggi (2018). Leading-by-example, leadership legitimacy and cooperation: evidence from sequential move public goods games. Working Paper.

5.1 Introduction

Leadership is important for economic development, resolution of social dilemmas, and management of the common resources (Chattopadhyay and Duflo 2004, Jones and Olken 2005, von Rueden et al. 2014, Ostrom 2009). By acting as role models, leaders can shape expectations of their constituency and induce them to engage in behaviors that are beneficial to the group as a whole. For instance, leaders who behave generously, do not accept bribes, and live by healthy ethical principles, could nurture follower's beliefs that others will do the same. On the contrary, corrupt and selfish leaders can favor the insurgency of unethical morale, inducing followers to behave accordingly and possibly eroding group prosperity and welfare. This phenomenon is commonly referred to as "leading by example".

An extensive body of experimental literature has studied the effects of "leading by example" on cooperation in a sequential public goods game (PGG) framework, where one subject in a group makes her contribution first as the leader and others contribute after observing the leader's contribution. Leaders can affect voluntary contributions of the followers through their own contribution. Yet the evidence on the effect of having a leader on cooperation in sequential move games is mixed (Figuieres, Masclet, and Willinger 2012, Gachter et al. 2010, Guth et al. 2007, Haigner and Wakolbinger 2010, Levati, Sutter, and van der Heijden 2007, Moxnes and van der Heijden 2003, Potters, Sefton, and Vesterlund 2007). Leading-by-example, where leaders contribute a large amount to the public account, is found to be able to improve the voluntary contributions of the followers (e.g. Guth et al. 2007, Levati, Sutter, and van der Heijden 2007). Yet, because followers can take advantage of their higher contributions, not all leaders choose to lead by example. When they failed to do so, they fail to improve cooperation (e.g. Haigner and Wakolbinger 2010, Potters, Sefton, and Vesterlund 2007, Rivas and Sutter 2011). Given the importance of a leader's role in sustaining cooperation, understanding leaders' "lead by example" behavior is important for understanding how leaders help to build and sustain cooperation.

One important factor that affects leaders' behavior and the effectiveness of leadership is leadership legitimacy. A legitimate leader could be more willing or feel more obligated to set an example to others and followers may be more willing to follow leaders selected through a legitimate procedure. In experimental settings, the leadership legitimacy is typically manipulated by the leader selection procedure. Researchers have tried to select leaders with various procedures, such as using real life leaders (Jack and Recalde 2015), volunteer leaders

(Dannenberg 2015, Haigner and Wakolbinger 2010, Rivas and Sutter 2011), or elected leaders (Levy et al. 2011, Brandts, Cooper, and Weber 2015, Reuben and Timko 2017).

Identifying how leadership legitimacy affects leading-by-example and cooperation behavior is not easy in the commonly used framework. First, when examining the effect of leading-by-example, most studies compare contributions in a sequential move game and contributions in simultaneous move game. But it cannot distinguish whether leaders' higher contributions are due to the sense of responsibility or due to the sequential move game structure where their behavior can be observed. Second, when studying the effect of leadership legitimacy, it is difficult to distinguish the effect of leadership legitimacy from the selection bias associated with the leader selection procedure, especially when the mechanism behind the selection bias is often not entirely clear to the researchers or hard to control for. Third, when studying leaders' behavior using sequential move public goods games, simultaneous move public goods games are not ideal comparison groups. The sequential move game structure differs from the simultaneous move game structure in two different ways: from a follower's perspective, observing leaders' contribution brings additional information about the behavior of others, which by itself could affect followers' contributions; and from a leader's perspective, contributing first and being observed brings the opportunities to act strategically (i.e. to lead by example), which could affect leaders' contribution decisions. Comparisons between sequential move PGGs and simultaneous move PGGs fail to distinguish these two different sources of impact, and often attribute the whole impact to "leading-by-example".

Bearing these problems in mind, in this study, we explore the effect of having a leader on group cooperation and particularly the effect of being a leader and leadership legitimacy on leader's behavior in sequential public goods games. We employ a novel experimental design to select all leaders with the same procedure and manipulate their perception about the leader positions and the perception of leadership legitimacy. We first use an "implicit election" procedure to select leaders in all sessions to control for the leader selection bias. Then, we manipulate the information that leaders receive about how leaders are selected to generate different perceptions about the leader positions and legitimacy of leader positions. Specifically, we have three leader treatments with different perceptions of leader positions. In the "unaware" leader treatment, leaders are not informed about being first movers and receive the same instruction interface as in the simultaneous move game Thus we single out the effect of followers observing leaders' contribution first by comparing contributions in the "unaware" leader treatment and contribution simultaneous move games. In the "appointed" leader

treatment, leaders are informed that they are selected as first-movers. Therefore, by comparing contributions in "appointed" leader treatments and contributions in the "unaware" treatment, we can distinguish the impact of leader's behavior on group cooperation and leaders' behavior. And in the "legitimate" leader treatment, leaders are informed that they are first-movers and they are selected based on the implicit election. By comparing contributions in the "legitimate" leader treatment and contributions in the "appointed" leader treatment, we can further identify the effect of leadership legitimacy on public goods provision.

We conducted the experiment with 272 college students in Beijing Normal University in 2015. We find that having a leader improves cooperation in the public goods game. The increase of group contributions is induced by the leader's contributions. Being a leader makes leaders increase their contributions, and followers reciprocate but to a slightly smaller extent and thus they harvest the gains from increased group contributions. The perception of leadership legitimacy does not have additional impacts on the leaders' contribution level and the group contribution level. However, we find evidence suggesting that leadership legitimacy influences how leaders make their contributions decisions and update their beliefs in the repeated public goods game. It seems that only "legitimate" leaders show the strategic use of "leading-by-example", not "appointed" leaders, whose higher contributions as leaders are simply a reaction to the sequential move game structure where their behavior can be observed by others.

This study contributes to the literature in several ways. First, our study aims to distinguish the effect of leaders' "leading-by-example" behavior from the effect of the observability of leaders' behavior in improving cooperation. Observing leaders' contributions before making contribution decisions alone could provide additional information to the followers and could thus change followers' beliefs about how much others would contribute and thus changes their contribution decisions. Moreover, leaders could strategically use their first-mover positions to influence follower and thus usually change (increase) their contributions. By using an "unaware" leader treatment, we create a game environment where leaders act as players in simultaneous games so that changes in public goods provisions are caused by followers' reactions. Second, our experiment tackles the selection bias problem in leader selection process, especially when generating leadership legitimacy. Experimental studies in the literature has used various ways to create leadership legitimacy but did not provide proper control groups. Using an "implicit election" procedure and manipulating the information received by the leaders, we aim to select all leaders through the same procedure and create differences in the

perceptions of leader positions and leadership legitimacy across treatments. Finally, unlike previous studies that mainly focus on the effect of leaders on group contributions or follower contributions, this study particularly focuses on the behavior of leaders. We also elicit leaders' beliefs on followers' contributions in an incentive-compatible way, so that we can study the contribution and belief dynamics of leaders. To our best knowledge, this study is the first study that explores this topic.

The rest of the papers is organized as follows: Section 5.2 describes our experimental design and procedure and the subject pool; Section 5.3 presents and analyzes the experimental results; and Section 5.4 concludes.

5.2 Experimental design and procedure

5.2.1 Experimental design

The experiment consists of five stages. Stage 1 is the "implicit election" stage where we select leaders and build up leader legitimacy perception for later stages. The core of our experiment is the public goods game. Stage 2-5 are all public goods games (PGGs) in different forms. Stage 2 is a strategy form public goods game. Stage 3 is a repeated simultaneous move PGG for 10 rounds. The first three stages are the same for everyone. We introduce our leader treatments in Stage 4 and 5. There are three leader treatments and a control group. The control group (S) just repeats Stage 3 twice and the three treatment groups contain a 10-round sequential move PGG in each stage. The three treatments are the "unaware (L1)", "appointed (L2)" and "legitimate (L3)" leader treatments that manipulate leaders' perception. Table 1 shows the structure of the experiment and we will describe our detailed design in the rest of this section.

Table 1. Experiment structure

Stage 1	Implicit election						
Stage 2	Strategy form PGG						
Stage 3	Simultaneous PGG						
	Simultaneous	Sequential PGG					
Stage 4	S	**L1**-F1	**L1**-F2	**L2**-F1	**L2**-F2	**L3**-F1	**L3**-F2
Stage 5	S	**L1**-F2	**L1**-F1	**L2**-F2	**L2**-F1	**L3**-F2	**L3**-F1

Note: S stands for simultaneous PGG; L1, L2 and L3 are the three leader treatments; F1 and F2 are two follower treatments. Leader treatments and follower treatments are independent from each other.

5.2.1.1 Implicit election

Stage 1 is a key element of our experimental design. The stage has two purposes: first, to build the perception of leadership legitimacy (to be manipulated in later stages); and second, to account for the selection bias associated with leader selection procedure. We call the leader selection procedure "implicit elections", because in this stage we do not explicitly call for an election, but select leaders based on participants' personality traits and their preferred traits of leaders, and the participants were not informed about the purpose of this stage yet.

Specifically, we use two personality tests, i.e. Machiavellianism Test and Authentic Leadership Test[51], for the implicit election mechanism. There were 20 questions in the Machiavellianism test and 16 questions in the Authentic Leadership test. Answers to the test questions are on five-point scale. In both tests, participants were asked first to finish each test for themselves and then to report the answers they preferred from a "qualified" leader. By eliciting participants' own personality traits and their preferences for the traits of leaders, we expect to create a sense of leadership legitimacy for the participants, when they are informed that leaders are selected based on their own traits and their preferences for leaders in later stages of the experiment.

The selection of leaders is based on participants' own answers and preferred answers in the Machiavellianism test. We use a 5-point scale answers to each question, assigning 1 point to the least Machiavellian answers and 5 points to the most Machiavellian answer. Adding up the scores of all answers to the 20 questions, we get the Machiavellianism scores ranging from 20 to 100, where higher scores represent stronger Machiavellianism attitudes. We classify people with scores no higher than 60 as low Machiavellianism (Low-Machs) and people with scores higher than 60 as high Machiavellianism (High-Machs).[52] Based on the participants' answers to the test questions, we calculate the Machiavellianism scores for each participant and thus classify them as either a High-Mach or a Low-Mach. Based on participants' preferred answers for qualified leaders, we calculate their preferred Machiavellianism scores and thus we know whether they prefer a High-Mach leader or a Low-Mach leader. We treat participants' preferred type of leaders as their "implicit votes" and use a majority rule to determine whether

[51] We use the Machiavellianism test following Christie and Geis (1970) and the Authentic Leadership test following Walumbwa et al. (2008). We choose the Machiavellianism test because the statements in the test have strong value orientations and we expected people would have diverse answers and strong preferences in terms of these statements. We use the Authentic Leadership test as a framing device to help strengthen the perception of leadership legitimacy.

[52] Sixty is the score if one chooses "neutral" for all the questions. We select leaders around the neutral score to avoid selecting leaders with extreme Machiavellian scores and to select leaders with scores closer to the "average" score.

an experimental session prefers High-Mach or Low-Mach leaders. If the session prefers High-Mach leaders, we first choose High-Mach participants with lowest scores as leaders (for later treatments). We typically select five leaders in one experiment session (four in a few cases).[53] In case there is less High-Machs in the session than the required number of leaders, we selected the low-Machs with the highest scores to fill in the vacancies. And if a session prefers Low-Mach leaders, then the process works under the same principle but in the opposite direction. By choosing High-Machs with lowest scores and Low-Machs with highest scores as leaders, we not only select leaders using the same procedure in all treatments, but also select leaders that are similar in terms of their Machiavellianism scores.[54]

Note that in this "implicit election" stage, participants are not informed about the implications of their answers to later stages yet. They will only learn about their roles as leaders and followers later in the sequential move PGGs.

5.2.1.2 Strategy form PGG and repeated simultaneous move PGG

Stage 2 and Stage 3 largely follow the experimental design in Fischbacher and Gachter (2010). Stage 2 is a one-shot strategy form PGG following the P-experiment in Fischbacher and Gachter (2010). Participants are randomly assigned to groups of four. Each participant is given an initial endowment of 20 tokens. They can decide how many tokens to keep with themselves and how many to contribute to a public project. The payoff function can be expressed as follows: $\pi_i = 20 - g_i + 0.4\sum_{j=1}^{4} g_j$, where g_i is participant i's contribution to the public project. The marginal payoff factor of the public project is 0.4, providing incentives of free-riding. In its strategy form, each participant is asked to make one "unconditional contribution" decision and 21 "conditional contribution" decisions corresponding to 21 possible rounded average contribution levels of other group members.

Stage 3 is a repeated simultaneous move PGG that largely resembles the C-experiment in Fischbacher and Gachter (2010) except that we use a pattern design instead of a stranger design. The game structure and payoff function are the same as that in Stage 2. Participants play for 10 rounds in fixed groups. In each round, participants are asked to make their contribution decisions and report their beliefs about the average contribution of other group members. We elicit participants' beliefs about the average contribution of others with the incentivized method

[53] We need one leader in every four participants due to our public goods game structure. A typical session has 20 participants and 16 in two cases.
[54] It is noteworthy that the detailed leader selection procedure described here was never revealed to the participants. We will describe how participants are informed when introducing the treatment later in this section.

of Fischbacher and Gachter (2010).[55] After making contributions and reporting beliefs, participants are informed about the total contributions from their group in that round and the accuracy of their beliefs before entering the next round until the 10th round. This stage serves as a baseline for within-subject comparison and as a framing device for the "unaware" leader treatment.

5.2.1.3 Sequential move PGGs with leader treatments

Stage 4 and 5 are our key treatment stages. In both stage, we have one control group where participants play the same 10-round simultaneous move PGG as in Stage 3 and three treatment groups where participants play 10-round sequential move PGGs. The sequential move PGGs have the same basic game structure and payoff function as the simultaneous move PGG in Stage 3. One participant in each group gets the role as the leader and others play as followers.[56] In each round, leaders contribute first knowing their contributions will be observed by the followers and followers make their contributions after observing leaders' contributions. Regarding belief elicitation, leaders report their beliefs about the average contribution of the three followers, while followers only report their beliefs about the average contributions of the other two followers. The incentivized method in belief elicitation is the same as that in Stage 3.

We have three leader treatments that manipulate leaders' perception of the leader position and leader legitimacy in the sequential move PGGs. The first treatment is "unaware" leader treatment (L1), where leaders are not informed about them being first-movers, and they receive the same instruction as that in the simultaneous move PGG in Stage 3. Since leaders are not aware that they are first-movers, we expect "unaware" leaders to act just like players simultaneous move PGGs. The difference between "unaware" leader treatment and the control group is the information that the followers receive. Therefore, the difference between contributions in the "unaware" leader treatment and contributions in the control group can be attributed to how followers respond to the additional information of leaders' contribution.

[55] Belief elicitation is incentivized by rewarding participants for the accuracy of their beliefs. If one's belief (integral only) exactly equals to the average contribution of others (rounded to the nearest integral), then she will be rewarded with three extra tokens. The reward is two tokens, if the gap between belief and actual average contribution of others is one point, and one tokens If the gap is two points, the reward is one point. No reward if the gap is larger than two points.

[56] Note that despite that we use terms such as "leaders" and "followers" in the text to facilitate writing, we did not use these terms in the experiment and only described the leader and followers as "contribute first" and "contribute after observe the first-movers' contributions".

The second treatment is "appointed" leader treatment (L2), where leaders are informed about them being selected as leaders, but not about the selection procedure. We name the treatment "appointed", since leaders are simply told that they are "selected" as leaders. We expect "appointed" leaders to consider their positions as exogenously determined and not attach legitimacy to their positions. As the only difference between the "unaware" leader treatment and the "appointed" leader treatment is leaders' awareness of their first-mover positions, and therefore, the difference between contributions in the "appointed" leader treatment and contributions in the "unaware" leader treatment must have stem from changes in leaders' behavior due to the leader positions.

The third leader treatment is "legitimate" leader treatment (L3), where leaders are informed that they are first-movers, and they are selected with based on their own answers and the session's preference of leaders in Stage 1, which is designed to raise a sense of legitimacy for the leader's role. By comparing "appointed" leader treatment and "legitimate" leader treatment, we are able to distinguish the effect of leadership legitimacy.

We use a between-subject design in Stage 4 and 5. Experimental sessions are randomly assigned to one of the three treatments or control groups. One participant can only play in one leader treatment. [57]

Leaders in all sequential move PGGs treatments are selected according to the "implicit election" procedure in Stage 1. Once selected, leaders will keep their role as leaders in both Stage 4 and Stage 5, while others keep playing as followers. Leaders and followers are randomly matched at the beginning of Stage 4 and 5, and then the leader-follower compositions remain constant over the 10 rounds.

5.2.2 Experimental procedure and subject pool

[57] The difference between Stage 4 and Stage 5 is the follower treatments. We manipulate followers' perception of leader legitimacy in a similar manner as we do in the leader treatments. In one follower treatment (F1), second movers are informed that one participant is selected as the leader in each group, but not the leader selection procedure. In the other follower treatment (F2), second movers are informed that there is a leader, and that the leader is chosen based on their answers and stated preferences in Stage 1. The follower treatments use a within-subject design. Participants play one treatment for 10 rounds and then in stage 4 and then the other for another 10 rounds. And half of the sessions played F1 first and the other half played F2 first as shown in Table 1. The leader treatments and the follower treatments are independent from each other. Leaders are not informed about the follower treatments and followers have no information about the leader treatments. For the leaders, there is no difference between Stage 4 and 5. Stage 5 is simply a repetition of Stage 4. This study focuses on leaders' behavior, and therefore we will not spend much time on the follower treatments.

The experiment was conducted with z-Tree (Fischbacher 2007) in the economics laboratory at Beijing Normal University in November 2015. Participants were randomly seated in the lab and separated by booths so that they could not see each other's decisions on the computer screens. Communication among participants was prohibited. Before the beginning of the experiment, participants were given a general instruction in paper (the original instruction was in Chinese, see Appendix A2 for the English translation). The general instruction explained the basic structure of the linear PGG along with a set of practice questions illustrating how the payoffs were calculated and the social dilemma situation in PGGs. Participants were informed that there would be different stages of PGGs, and the PGG in one stage could be the same as or different from the PGG in the previous stage and they would get instructions on the computer screens in each stage of the experiment.

We recruited experiment participants through the university's Bulletin Board System. A total of 272 students participated in 14 experiment sessions. About 73% of the participants were female which loosely represented the gender ratio of the university. Three quarters of the participants were undergraduate students. A large majority (73%) of the participants had no prior experience of economics experiments and only a small proportion (8%) had learnt game theory before the experiment.

Experiment sessions were randomly assigned to the three leader treatments and the control group. Four sessions were assigned to each of the three leader treatments, and the rest two sessions were assigned to the control group. A typical experimental session consisted of 20 participants, with two exceptions having 16 participants. There were 76 participants in treatment L1, 80 in L2 and 80 in L3, while the rest 36 participants were assigned to the control group. Therefore, there were 19 "unaware" leaders in L1, 20 "appointed" leaders in L2 and 20 "legitimate" leaders in L3.

5.3 Experimental Results

5.3.1 Leader selection

One key feature of our experimental design is that we select leaders using the same procedure across treatments to avoid potential selection biases. Therefore, we first look at who are selected as leaders for later sequential move PGGs and whether the leaders selected in different leader treatments are indeed similar in terms of personal characteristics and especially answers in their personality tests. Table 2 shows summary statistics of the relevant

characteristics of the selected leaders in three treatments. There are not much differences in age, gender, authentic leadership scores and preferred authentic leadership scores of the leaders among the three treatments.[58] Leader's Machiavellianism test scores as well as the scores they prefer fall in similar range across the three treatments.[59] We believe that our leader selection procedure select similar leaders for all treatment and selection bias would not be an serious issue in our effort to identify the impact of being a leader and of the leadership legitimacy on cooperation in PGGs in this study.

Machiavellianism test scores of all leaders selected from the "implicit election" procedure fall into the range from 48 to 59. This means two things. First, they are all Low-Machs, whose scores are under 60. This is because it is few participants in our sample score 60 or higher, although relatively higher scores are preferred for leaders. [60] Second, the Machiavellianism test scores of the whole sample range from 25 to 74 with a median of 45. This means that we only select leaders from the upper half of the score distribution. They are not represented of the whole sample, but close to what an average participant prefers as leaders (See Appendix A1).

[58] For gender, joint Fisther's exact=0.930, pair-wise Fisther's exact>0,716; for age, Kruskal-Wallis test, p=0.733, pair-wise M-W tests, p>0.473; for authentic leadership scores, Kruskal-Wallis test, p=0.480, pair-wise M-W tests, p>0.230; for preferred authentic leadership scores, Kruskal-Wallis test, p=0.480, pair-wise M-W tests, p>0.131.
[59] Fishers' exact test (Fisher's exact = 0.508) suggests that there is no joint statistical difference in the Machiavellianism scores' distribution across treatments Yet, we cannot reject the possibility of differences in means among the three groups if we treat MT scores as a continuous variable. However, there is no significant correlation between leaders' MT scores and leaders' contributions in PGG either in each treatment or in pooled leader sample and the correlations between leaders' MT scores and leaders' contributions within each treatment are not significantly different from each other.
[60] Appendix A1 provides the distribution of Machiavellianism test score and preferred Machiavellianism test scores for leaders.

Table 2. Leaders selected through the across leader treatments

		L1, “unaware”	L2, “appointed”	L3, “legitimate”
	No.	19	20	20
Male	No.	5	5	4
Age	Mean	20.7	19.9	20.2
	Min	17	17	17
	Max	26	25	26
Machiavellian Test Score	Mean	51.6	52	54.4
	Min	48	48	50
	Max	57	59	59
Authentic Leadership Score	Mean	56.4	56.2	55.7
	Min	39	39	43
	Max	68	66	70
Preferred Machiavellian Test Score	Mean	55.7	57.1	56.6
	Min	42	41	40
	Max	69	72	67
Preferred Authentic Leadership Score	Mean	67.1	64.4	63.4
	Min	52	54	50
	Max	80	75	75

Note: Calculated by the authors.

Table 3 shows participants’ average contributions and the standard deviations of the contributions in different treatment or control sessions in different stages of the experiment. Figure 1, 2 and 3 show the how the average contributions of all group member, leaders and followers evolve over three stages of 10-round PGGs. In these figures, round 1-10, round 11-20 and round 21-30 correspond to Stage 3, Stage 4 and Stage 5 respectively. We will describe and analyze these results in details in the rest of this section.

Table 3. Average contributions in each stage by treatments

		N	Stage 3: Round 1-10 Mean (SD)	Stage 4: Round 11-20 Mean (SD)	Stage 5: Round 21-30 Mean (SD)
S	Group	36	5.67 (4.87)	4.13 (4.34)	3.59 (3.82)
L1	Group	76	5.17 (5.56)	2.83 (4.55)	3.18 (4.62)
	Leaders	19	5.74 (6.50)	3.46 (5.67)	3.74 (5.45)
	Followers	57	4.98 (5.40)	2.62 (4.10)	2.99 (4.30)
L2	Group	80	4.62 (4.93)	5.13 (6.25)	5.51 (7.04)
	Leaders	20	5.46 (6.01)	6.08 (6.94)	6.61 (7.27)
	Followers	60	4.35 (4.48)	4.82 (5.98)	5.15 (6.93)
L3	Group	80	3.71 (4.20)	4.28 (6.18)	3.02 (5.00)
	Leaders	20	3.77 (4.27)	5.04 (6.55)	3.44 (4.81)
	Followers	60	3.69 (4.18)	4.03 (6.04)	2.89 (5.05)

Note: calculated by the authors.

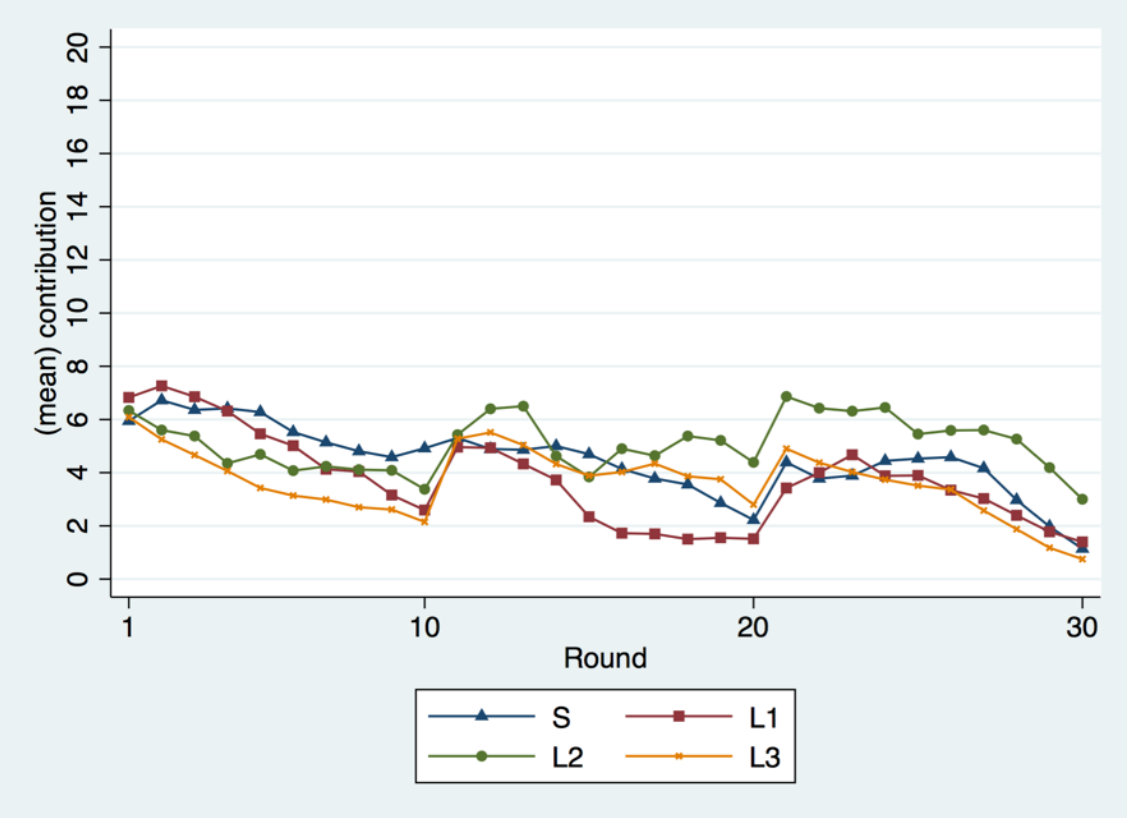


Figure 1. Evolution of average contributions in control and treatment sessions
The blue line with triangle marks shows contributions of subjects in the control group (S); the red line with squared marks shows contributions of subjects in the "unaware" leader treatment (L1); the green line with circle marks shows contributions of subjects in the "appointed" leader treatment (L2); the orange line with x marks shows contributions of subjects in the "legitimate" leader treatment (L3).

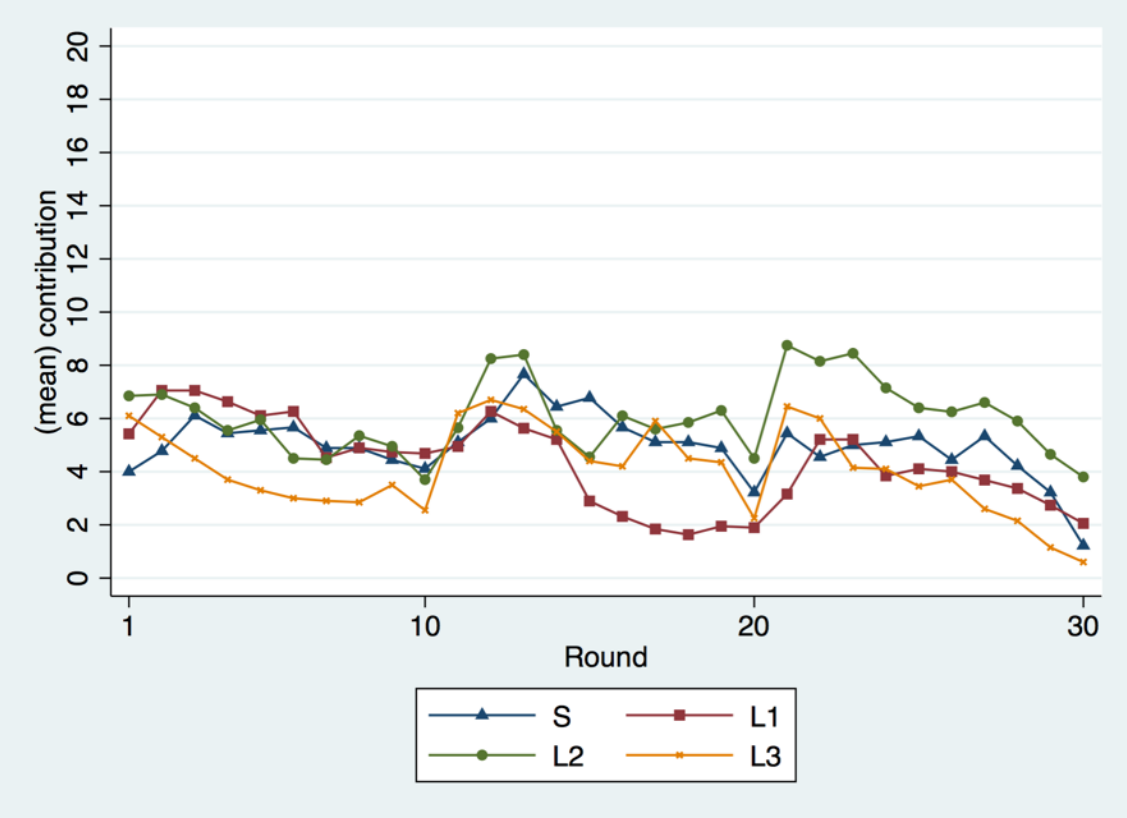


Figure 2. Evolution of leaders' contributions in control and treatment sessions
The blue line with triangle marks shows contributions of leader candidates selected through the implicit election in the control group (S); the red line with squared marks shows contributions of leaders in the "unaware" leader treatment (L1); the green line with circle marks shows contributions of leaders in the "appointed" leader treatment (L2); the orange line with x marks shows contributions of leaders in the "legitimate" leader treatment (L3).

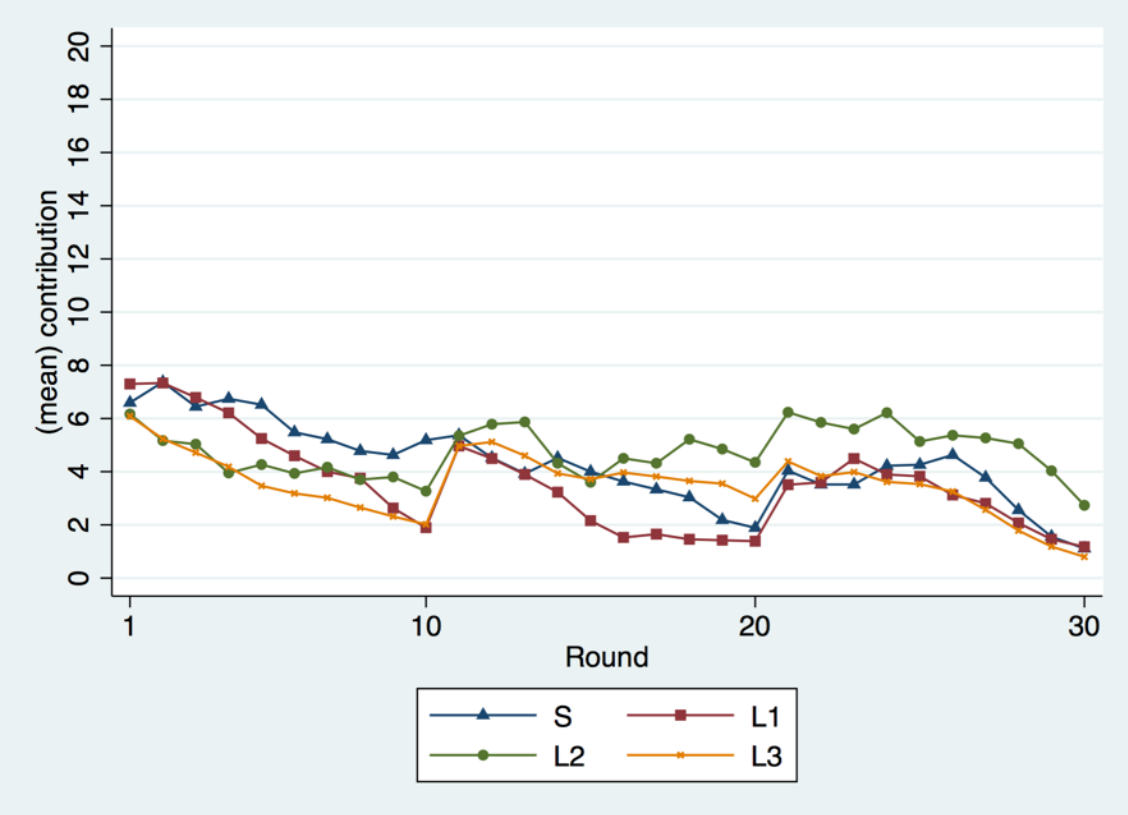


Figure 3: Evolution of followers' contributions in control and treatment sessions
The blue line with triangle marks shows contributions of follower candidates selected through the implicit election in the control group (S); the red line with squared marks shows contributions of followers in the "unaware" leader treatment (L1); the green line with circle marks shows contributions of followers in the "appointed" leader treatment (L2); the orange line with x marks shows contributions of followers in the "legitimate" leader treatment (L3).

5.3.2 Declining contributions in repeated simultaneous move public goods games

We first look at contributions in Stage 3 where all participants play simultaneous move PGGs. Round 1-10 in Figure 1 shows how average contributions evolve in Stage 3. The average contribution of all participants in first round is 6.35 tokens, about 32% of the initial endowment. There is a declining trend of cooperation over the 10 rounds, similar to the findings in literature on repeated PGGs, e.g. Fischbacher and Gachter (2010) and Gachter and Renner (2010). At the end of Stage 3, the average contribution drops to 3 tokens, about half the size of contributions in the first round.

We expect participants to behave similarly in this stage regardless which control and treatment group they are assigned to later in the experiment. Figure 1 to Figure 3 shows some mean differences in contributions over the 10 rounds across treatments. Participants in control sessions (S) contribute a bit more and participants in "legitimate" leader treatment (L3) contribute a bit less. We test the differences in contributions across control and treatment

sessions both jointly and pair-wise, and we cannot reject the null hypothesis of no differences in contributions (Kruskal-Wallis test, $p=0.468$; Mann-Whitney U test, $p>0.141$). We also test this in by regressing treatment dummies on contributions and results are shown in column (1) in Table 4. The coefficients of three leader treatment dummies are not significantly different from zero, which confirms the aforementioned non-parametric test results. In column (2), we add interaction terms between treatment dummies and game rounds as additional explanatory variables find that there are also no significant differences in the decline trends across treatments.

Since we select all leaders from the same procedure, we also expect that both leaders and followers across control and treatment sessions act similarly in Stage 3. Round 1-10 in Figure 2 and Figure 3 are the contribution evolution of participants who are selected as leaders and followers for later stages, respectively. We find no statistically differences in the contributions of leaders (Kruskal-Wallis test, $p=0.802$; pair-wise Mann-Whitney U test, $p>0.444$) and followers (Kruskal-Wallis test, $p=0.457$; pair-wise Mann-Whitney U test, $p>0.126$). We also test this in column (3) and (4) in Table 4, and none of the coefficients of treatment dummies are statistically significant. However, despite the statistically insignificant, the size of the coefficients of treatment dummies (especially for L2 and L3) in Table 4 is quite large. We believe it is better to take these differences into consideration when we evaluate the effects of our leader treatments on group cooperation and leader behavior later in this paper.

Note that our leader selection procedure only selects leaders whose Machiavellianism scores lie within a certain range (i.e. 48-59, see Table 2). Thus, the selected leaders are different from the followers in terms of Machiavellianism scores and thus leaders may also be different from followers in terms of contributions in PGGs. Column (5) in Table 4 tests whether there is a difference in contributions between potential leaders and followers in Stage 3 and we find that our leaders selection procedure selects more cooperative participants as leaders for later sequential move PGGs. We bare this in mind when comparing the contributions of leader with followers in sequential move PGGs.

Table 4. Balance check of contributions in simultaneous move PGG

	(1)	(2)	(3)	(4)	(5)
Contribution	Pooled	Pooled	Leader	Follower	Pooled
Round	-0.377***	-0.222	-0.266***	-0.415***	-0.377***
	(0.0483)	(0.236)	(0.0647)	(0.0522)	(0.0483)
L1	-0.362	1.350	0.863	-0.760	-0.365
	(1.234)	(1.353)	(1.907)	(1.349)	(1.209)
L2	-1.079	-0.815	0.107	-1.579	-1.055
	(1.157)	(1.172)	(1.814)	(1.249)	(1.127)
L3	-1.883	-0.861	-2.017	-2.046	-1.808
	(1.130)	(1.214)	(1.978)	(1.253)	(1.109)
Leader					1.214**
					(0.529)
MT score	-0.0309	-0.0309	0.232	-0.0951***	-0.0750**
	(0.0356)	(0.0357)	(0.232)	(0.0320)	(0.0320)
Age	-0.101	-0.101	-0.161	-0.0897	-0.104
	(0.0731)	(0.0731)	(0.189)	(0.0774)	(0.0706)
Experience	0.163	0.163	-0.824	0.576	0.174
	(0.473)	(0.474)	(1.073)	(0.528)	(0.470)
Game theory	-0.843	-0.843	-0.678	-0.827	-0.830
	(0.650)	(0.651)	(1.513)	(0.915)	(0.633)
Risk attitude	0.155	0.155	-0.0159	0.235*	0.167
	(0.141)	(0.141)	(0.215)	(0.136)	(0.142)
L1 * Period		-0.311			
		(0.254)			
L2 * Period		-0.0481			
		(0.243)			
L3 * Period		-0.186			
		(0.241)			
Constant	10.34***	9.485***	-1.804	12.64***	12.00***
	(2.530)	(2.691)	(10.73)	(2.814)	(2.415)
Observations	2720	2720	680	2040	2720
Adjusted R2	0.075	0.079	0.060	0.111	0.082

Note: OLS regression results for contributions in Stage 3 are in the table. Robust standard errors clustered at group level are reported in the parentheses. Column (3) and (4) use the sub-sample of potential leaders and follower selected through the implicit election procedure in Stage 1. ***, ** and * stand for significant level 1%, 5% and 10%, respectively.

5.3.3 Treatment effects on group contributions

Now we move to analyze the effects of different treatments on the average contribution at group level. Group contributions to the public goods determines the social efficiency in the PGG setting. Table 3 shows the average contributions in each treatment and Figure 1 shows how contributions in different treatments evolve in Stage 4 and 5. We see that average group contributions have a jump start at the beginning of either stage and decline over the 10 rounds of either stage in all control and treatment groups.

Because there exist some differences (despite statistically insignificant) in participants' contributions across control and treatment sessions in Stage 3, these prior differences could be carried over to later stages and obscure the treatments effects if we only focus on between-subject comparison in Stage 4 and 5. Therefore, when analyzing treatment effects, we use

difference-in-difference regressions that also take participants' behavior in Stage 3 into account and control for individual fixed effects. First, we compare all three leader treatments (L1/L2/L3) with control groups (S). The regression results are in Table 5.

Column (1) in Table 5 shows the regression results for the effects of different sequential move PGG treatments on contributions using simultaneous move PGG as control group. We find that having "unaware" leaders does not increase group contributions, while having "appointed" leaders and "legitimate" leader increase cooperation, comparing with PGG with no leaders. However, groups with "legitimate" leaders do not outperform groups with "appointed" leaders. And naturally, because contributions increase in L2 and L3, participants in these two treatments enjoy higher payoffs per round, which is shown in column (2) in Table 5.

Table 5. Treatment effects on group contributions

	(1) Contribution	(2) Payoff
L1: "Unaware"	-0.351	-0.210
	(0.669)	(0.402)
L2 "Appointed"	2.509***	1.506***
	(0.794)	(0.476)
L3: "Legitimate"	1.759***	1.055***
	(0.493)	(0.296)
Period	-0.338***	-0.203***
	(0.0402)	(0.0241)
Stage 4	-1.697***	-1.018***
	(0.359)	(0.215)
Stage 5	-1.926***	-1.155***
	(0.286)	(0.172)
Constant	6.504***	23.90***
	(0.346)	(0.208)
Observations	8160	8160
Adjusted R2	0.431	0.263

Note: OLS regression results. Individual fixed effects are omitted from the table for brevity. Robust standard errors clustered at session level are reported in the parentheses. Payoff here is the payoff from PGG in each round, excluding the payoff from belief elicitation. ***, ** and * stand for significant level 1%, 5% and 10%, respectively.

Table 5 compares contributions in each sequential move PGG treatment with the simultaneous move PGG control group to identify the overall effect of each treatment. But this study is not only interested in the overall effects, but also aims to distinguish the effects of different factors, i.e. followers' response to the knowledge about leaders' contributions, leaders' response to leader positions and the leadership legitimacy. For this purpose, we compare contributions in different treatment pair-wisely. Specially, we compare the control group (S)

with the “unaware” leader treatment (L1) to identify the effects of followers’ response to the knowledge about leaders’ contributions; compare the “unaware” leader treatment (L1) with the “appointed” leader treatment (L2) to identify the effect of leaders’ response to their leader positions and compare “appointed” leader treatment (L2) with “legitimate” leader treatment (L3) to identify the effect of leadership legitimacy. The regression results are in Table 6.

Column (1) and (2) of Table 6 compare the average contributions in control group (S) and “unaware” leader treatment (L1). Column (1) treats “unaware” leader treatment in Stage 4 and 5 as a whole and column (2) includes “unaware” leader treatment in Stage 4 and Stage 5 separately. Since the leaders in L1 are “unaware” of their positions, all differences in average contributions come from the followers’ reaction to the knowledge of leader’s contributions before making their own contribution decisions. We find that there is no significant difference between average contributions in sequential move PGGs with “unaware” leaders and contributions in simultaneous move PGGs. Followers’ responses to knowing leaders’ contributions by itself does not increase pubic good provision at group level. If anything, contributions suffer a subtle decline in the “unaware” leader treatment (L1).

Table 6. Treatment effects on average group contributions

	(1) Contribution : L1 vs. S	(2) Contribution : L1 vs. S	(3) Contribution : L2 vs. L1	(4) Contribution : L2 vs. L1	(5) Contribution : L3 vs. L2	(6) Contribution : L3 vs. L2
"Unaware"	-0.351					
	(0.707)					
"Unaware"		-0.799				
Stage 4		(0.905)				
"Unaware"		0.0978				
Stage 5		(0.859)				
"Appointed"			2.860**			
			(1.016)			
"Appointed"				2.844**		
Stage 4				(1.033)		
"Appointed"				2.876*		
Stage 5				(1.354)		
"Legitimate"					-0.751	
					(0.903)	
"Legitimate"						0.0675
Stage 4						(1.076)
"Legitimate"						-1.569
Stage 5						(1.124)
Period	-0.379***	-0.379***	-0.338***	-0.338***	-0.310***	-0.310***
	(0.0898)	(0.0898)	(0.0675)	(0.0675)	(0.0290)	(0.0290)
Stage 4	-1.843***	-1.539*	-2.346**	-2.338***	0.915	0.506
	(0.438)	(0.662)	(0.672)	(0.602)	(0.789)	(0.840)
Stage 5	-1.779***	-2.083***	-1.978**	-1.986**	0.481	0.890
	(0.264)	(0.178)	(0.776)	(0.820)	(0.922)	(1.077)
Ind. fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Constant	7.412***	7.412***	6.745***	6.745***	5.868***	5.868***
	(0.743)	(0.743)	(0.600)	(0.601)	(0.315)	(0.315)
Observations	3360	3360	4680	4680	4800	4800
Adj. R2	0.437	0.438	0.431	0.431	0.428	0.432

Note: OLS regression results. Robust standard errors clustered at session level are reported in the parentheses. Payoff here is the payoff from PGG in each round, excluding the payoff from belief elicitation. ***, ** and * stand for significant level 1%, 5% and 10%, respectively.

Then, we look at how leaders' response to their leader positions affects group contributions. Column (3) and (4) compare contribution in the "unaware" leader treatment (L1) and in the "appointed" leader treatment (L2). The only difference between these two treatments is whether the leaders are aware of their leader positions or not, and thus we can attribute the differences in contribution to the effect of leaders' response to leader positions. We find that having leaders' responses to their positions significantly increase group contributions. Combining the results from the first two column, this result means that the increase in group contributions found in having an exogenously appointed leader (coefficient of L2 in Table 5) is entirely induced by the changes in leader's behavior. As long as leaders do not change their behavior, followers' response to the knowledge about leader's behavior alone does not increase

their contributions. The increase in group contributions is also persistent throughout Stage 4 and 5, as shown in column (4).

Finally, we examine whether the leader's perception of the legitimacy of their leader positions affect group contribution, by comparing "appointed" leader treatment (L2) and in "legitimate" leader treatment (L3). The results are shown in column (5) and (6) in Table 6. Surprisingly, we find that groups with "legitimate" leaders contribute no more, if not less, than groups with "appointed" leader. Leader's perception about leadership legitimacy does not further increase group contributions.

5.3.4 Contributions of leaders and followers

We have found that sequential move PGG with leaders improve cooperation at the group level. The increase in group contributions may come from two sources: changes in the leaders' contributions and the increase in followers' contributions. Leaders may increase their contribution to induce higher contributions from the followers, while followers may reciprocate the leaders or take advantages of them. Therefore, we further separate leaders and followers in sequential move PGGs as different treatment dummies to see to whether the increase in average contribution found in previous sub-section is from the leaders or the followers. Table 7 shows the regression results using the same DID framework.

Column (1) – (3) of Table 7 shows the effects of leader treatments on both leaders' and followers' contributions. First, contributions of "unaware" leaders do not differ much (slight lowers, but not significant) from contributions in simultaneous PGGs. This is consistent with our expectations as "unaware" leaders received the same instruction in "unaware" leader treatment in Stage 4 and 5 as what they received in simultaneous move PGGs in Stage 3. Second, both "appointed" and "legitimate" leaders and followers increase their contribution, but neither leaders nor follower in "legitimate" leader treatment are doing any better than those in "appointed" leader treatment (coefficients of "legitimate" treatments are actual smaller).

Besides the effectiveness of different leader treatments in improving cooperation, we are also interested in comparing effects on leaders with the effects on followers. When having leaders has positive effects on group contributions, the relative size of the effects on leaders and followers affects how the gains from higher public goods provision are distributed. Comparing the coefficients of leader dummies and follower dummies, we find that effects on leaders are slight larger in all three treatments (e.g. in column (1), -0.328 vs. -0.358 in L1,

2.694 vs. 2.448 in L2 and 2.276 vs. 1.586 in L3). Although these differences are not statistically significant, remember that when it comes to payoffs, there are three followers matching one leader, meaning that the small differences in contribution between leaders and followers individually could add up to larger differences in pay offers. We examine how effects on the payoffs are distributed in column (4) of Table 7. We can see that, in "unaware" leader treatment, although there is no efficiency gain or loss at group level (column (2) in Table 5), "unaware" leaders actually suffer a loss, while followers' payoffs are not affected, if not higher. And in both the "appointed" and the "legitimate" leader treatments, despite the fact that group welfare increases due to increased contributions, it is the followers, rather than the leaders, that capture the efficiency gain. Leaders' payoffs barely change despite their higher contributions to the public goods. These results confirm the existence of the "leaders' curse" (Gächter and Renner 2014), where followers take advantage of the leaders' higher contributions. When leaders do not change their contributions (as in L1 where they are not aware of being first-movers), followers contribute slightly less, which hurts the leaders. And when leaders increase their contributions as a response to their leader positions, followers only partially reciprocate so that they benefit more than the leaders.

Table 7. Treatment effects on leaders' and followers' contributions

	(1) Pooled Contribution	(2) Stage 3 & 4 Contribution	(3) Stage 3 & 5 Contribution	(4) Pooled Payoff
Leader: "Unaware"	-0.328	-0.740	0.0833	-1.709**
	(0.724)	(0.939)	(0.777)	(0.581)
"Appointed"	2.694**	2.154	3.233***	0.295
	(0.919)	(1.414)	(0.966)	(0.780)
"Legitimate"	2.276**	2.804***	1.748*	0.0157
	(0.772)	(0.908)	(0.911)	(0.387)
Follower: "Unaware"	-0.358	-0.819	0.103	0.289
	(0.672)	(0.869)	(0.848)	(0.417)
"Appointed"	2.448***	2.009*	2.887**	1.909**
	(0.807)	(0.956)	(1.129)	(0.691)
"Legitimate"	1.586***	1.882*	1.290***	1.402***
	(0.451)	(0.940)	(0.225)	(0.363)
Period	-0.338***	-0.330***	-0.366***	-0.203***
	(0.0402)	(0.0500)	(0.0484)	(0.0241)
Stage=4	-1.697***	-1.539**		-1.018***
	(0.359)	(0.633)		(0.215)
Stage=5	-1.926***		-2.083***	-1.155***
	(0.286)		(0.170)	(0.172)
Constant	6.504***	6.456***	6.659***	23.90***
	(0.346)	(0.379)	(0.374)	(0.208)
Observations	8160	5440	5440	8160
Adjusted R2	0.431	0.472	0.466	0.269

Note: Robust standard errors clustered at session level are reported in the parentheses. Payoff here is the payoff from PGG in each round, excluding the payoff from belief elicitation. ***, ** and * stand for significant level 1%, 5% and 10%, respectively.

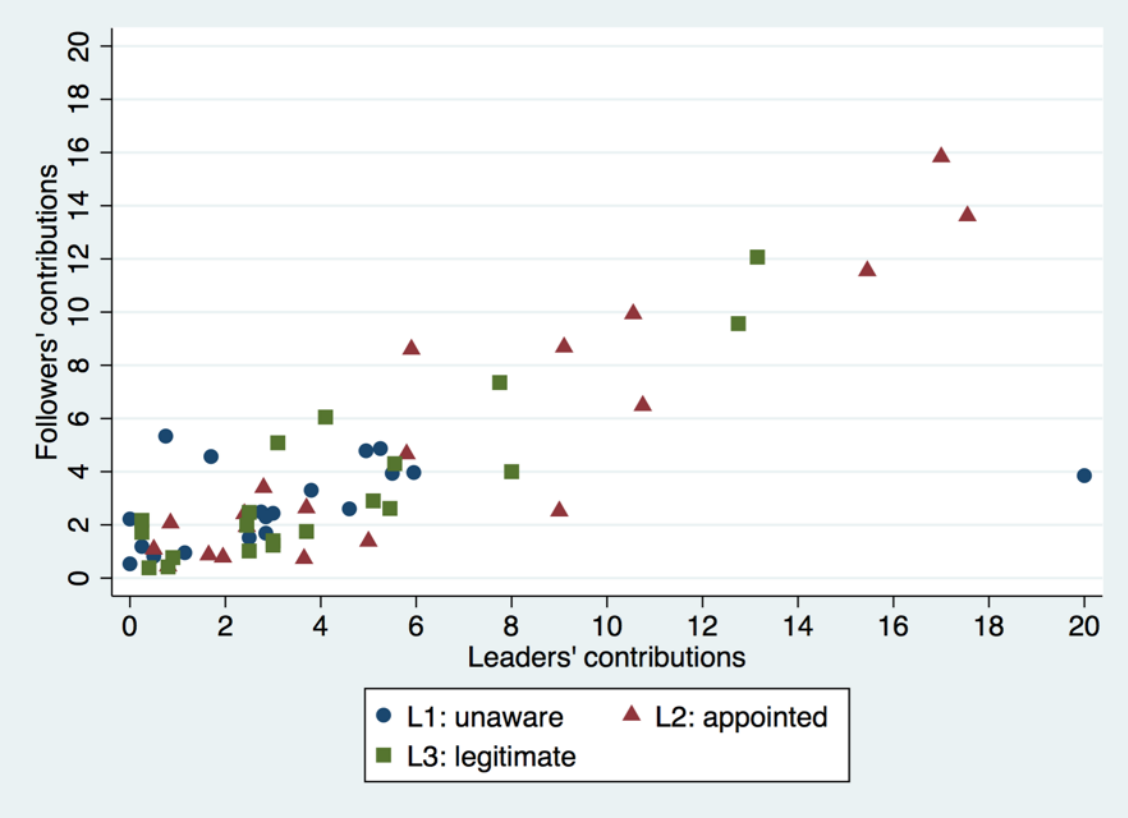


Figure 4. Leaders' contributions and followers' contributions in sequential move PGG

So far, we have found that leaders, or leaders' contributions, are crucial in sequential move PGGs. Followers response to the knowledge of leaders' contribution *per se* does not lead to better cooperation. When the leaders act as players in a simultaneous game and do not know their leader positions (as in L1), we do not find significant increase in contribution. Only in treatments where leaders actively increase their contributions as a response to their leader positions (as in L2 and L3), followers' contributions increase and the groups enjoy efficiency gains. Leaders' contribution level determines the effectiveness of sequential move game structure. Figure 4 depicts the relationship between leaders' contributions and followers' contribution. Followers respond to leaders' contributions in an imperfect conditional cooperator manner, and this relationship is valid both within-treatment and across treatment. The "appointed" leader treatment and the "legitimate" leader treatment are effective in promoting cooperation only because they change leaders' behavior.

However, despite that leaders increase their contributions when acting as leaders, they do not benefit more in terms of payoffs. Thus, we are interested in what motivated leaders to increase their contribution. We expected that leadership legitimacy to play a role. Yet, despite

the fact that our design overcomes the selection bias usually related to leadership legitimacy, it is still surprising to find that leadership legitimacy does not increase leaders' contributions.

Why do leaders not respond to the legitimacy of their leader positions? One possible reason is that the perception of leadership legitimacy does not only affect leaders' contributions, but also their beliefs on the contributions of the followers. Studies have shown that beliefs are important in shaping peoples contributions in PGG settings (Fischbacher and Gachter 2010, Gachter and Renner 2010). If leaders form their beliefs differently, the implication of leadership legitimacy could be very different. Therefore, we continue to explore the interaction between beliefs and contributions in the sequential move PGGs.

5.3.5 Coevolution of contribution and beliefs

Figure 5 shows the evolution of average contributions and beliefs about the contributions of others in the sequential move PGGs by treatments and the roles in sequential PGG. Panels on the left are for the followers and panels on the right are for the leaders.

The contributions of followers are always slightly lower than their beliefs about the average contribution of others followers. The same tendency has been found in simultaneous move PGGs and Fischbacher and Gachter (2010) regard it as incomplete reciprocity and the reason of declining cooperation.

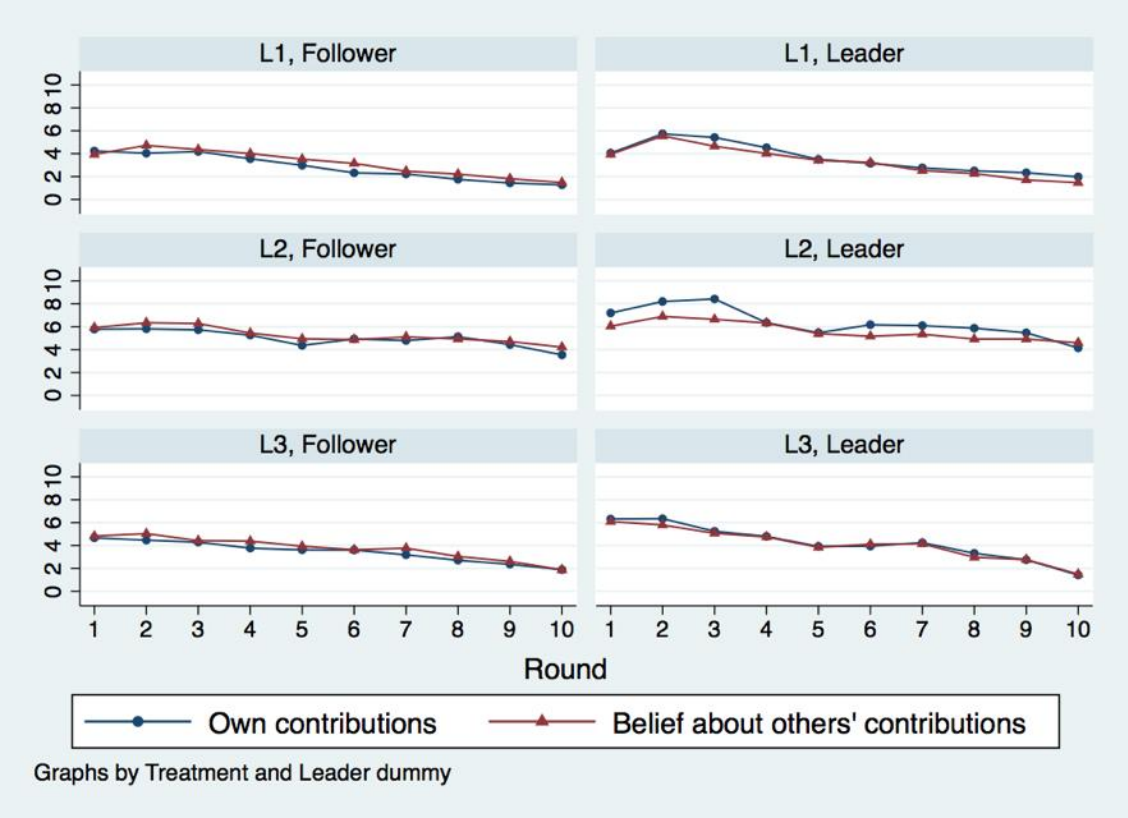


Figure 5. Contribution and belief dynamics in sequential move PGGs
L1, L2 and L3 stand for the "unaware" leader, "appointed" leader and "legitimate" leader treatment respectively.

For leaders, "appointed" leaders' contributions are constantly higher than their belief about the contribution of the followers, while "legitimate" leaders' contributions are very close to their beliefs of others' contributions, and more similar to "unaware" leaders, except that "legitimate" leaders started with higher contribution in the first round. Since the only difference between the "legitimate" leader treatment and the "appointed" leader treatment is that "legitimate" leaders are informed that they are selected based on the implicit election, the sense of legitimacy may have affected the belief formation dynamics so that the belief formation of "legitimate" leader follows a different trajectory of that of "appointed" leaders. Therefore, we continue to look into dynamics of contribution and belief of the leaders.

5.3.5.1 Dynamics of leaders' contribution decisions

First, we look at how the leaders update their contributions in the repeated PGGs. Following contribution and belief dynamics analysis in Fischbacher and Gachter (2010) and Gächter and Renner (2014), we assume that leaders adjusted their contributions based on their

beliefs about the contributions of followers and their contributions in the previous round. A larger coefficient of the belief variable implies that participants are closer to be conditional cooperators and perfect conditional cooperators when the coefficient equals one.

Including the lagged dependent variable as an explanatory variable is very likely to introduce endogeneity and it often overestimates the coefficient of the lagged dependent variable and beliefs could also be endogenous to contribution decision due to "projection bias" or "false consensus effect" (Ross, Greene, and House 1977), we use the Blundell-Bond system GMM estimation for dynamic panel data (Blundell and Bonds 1998) to estimate the dynamics of contribution decisions.[61] Results are shown in Table 8. Besides the leaders in the three leader treatments, we also use the people identified with the same "implicit election" procedure in the control group as additional "potential leader" control.

Table 8. Contribution dynamics of leaders, by treatment

Contribution	(1) S "potential leader"	(2) L1 "unaware leader"	(3) L2 "appointed leader"	(4) L3 "legitimate leader"
Contribution (t-1)	0.383***	0.417***	0.152	0.0188
	(0.0778)	(0.117)	(0.0934)	(0.0452)
Belief	0.513***	0.496***	0.698***	0.839***
	(0.0936)	(0.182)	(0.0778)	(0.116)
Constant	-0.178***	-0.0900*	-0.148	-0.0987
	(0.0657)	(0.0473)	(0.114)	(0.0601)
Observations	162	342	360	360

Note: Blundell-Bond estimation results. Robust standard errors clustered at session level are reported in the parentheses. ***, ** and * stand for significant level 1%, 5% and 10%, respectively.

First, we find that the sum of the coefficient of lagged contribution and belief is not significantly different from 1 in all treatments, suggesting that leaders' contributions are weighted averages of their contributions in the previous round and their beliefs of followers' contributions. Second, we find that the weight "unaware" leaders put on beliefs in L1 does not differ much from the weight that potential leaders put on beliefs in simultaneous PGG in S. This confirms our expectations that "unaware" leaders in L1 should behave like those in simultaneous PGG since they receive the same information. The interesting finding is that coefficients of beliefs in L2 is larger than that in L1 and the coefficient of beliefs in L3 is larger than that in L2. This implies that when aware of being leaders, people act differently:

[61] We also followed the procedure proposed by Smith (2013) to deal with the potential endogeneity of beliefs and the results are similar.

knowingly acting as leaders, people tend to behave more like conditional cooperators and more incline to match their contributions with their expectations about the average contribution of others. and knowing their "legitimacy" as first movers seems to further increase the conditional cooperation tendency.

5.3.5.2 Dynamics of leaders' belief updating

As we find that leaders act more like conditional cooperators that match their contribution with their beliefs about others, it is thus important to understand how leaders form and update their beliefs about followers in repeated PGGs. As we elicit the beliefs in an incentive-compatible way, we assume that all participants are trying to form accurate beliefs. In Fischbacher and Gachter (2010) and Gächter and Renner (2014), beliefs about the others' contribution is assumed as a function of one's belief in the previous round and the observed average contribution of others in the previous round. However, different from the participants in simultaneous move PGGs as in Fischbacher and Gachter (2010) and followers in sequential move PGGs as in Gächter and Renner (2014), this study focuses on the belief formation of leaders. In the positions as first movers, leaders could use their positions strategically, which is, to lead by example. The term "leading by example" has two-fold meanings: the leaders contribute more to set an "example" to others to induce better cooperation and the leaders need to "lead" knowingly, meaning that the leaders would expect their higher contribution could influence the followers' contributions. Thus, if leaders are trying to "lead", they would anticipate that their own contributions affect followers' contributions and take this into account when forming their beliefs about the contribution of followers. It means that if leaders try to "lead by example", their own contributions should affect their beliefs about the average contributions of the followers. Thus, we treat leaders' belief formation process as a function of their beliefs in the previous round, the observed average contribution of others in the previous round and leaders' contributions in that round.

As in the contribution decision analysis, since all the explanatory variables are potentially endogenous, we use Blundell-Bond system GMM estimation and the treatment by treatment estimation results are reported in Table 9. First, we find that even "unaware" leaders' own contribution to some extent affect their beliefs on the followers' contribution. This is probably due to the "projection bias" or the false "consensus effect" where people project their own behavior to their prediction of other's behavior. Smith (2013) finds that beliefs are indeed

endogenous to contribution decisions. We run the same regression with belief formations in simultaneous games and we find that players in simultaneous PGGs have very similar "projection bias" pattern as "unaware" leaders. The "potential leaders" in the control group also show a similar pattern. As "unaware" leaders are supposed to behave as players in simultaneous PGGs, we believe that they can be a good comparison group for "appointed" and "legitimate" leaders.

Table 9. Belief updating dynamics of leaders, by treatment

Belief	(1) S "potential leader"	(2) L1 "unaware leader"	(3) L2 "appointed leader"	(4) L3 "legitimate leader"
Belief (t-1)	0.271***	0.0818***	0.127***	0.0730
	(0.0936)	(0.0266)	(0.0126)	(0.0741)
Others' contribution (t-1)	0.431***	0.617***	0.618***	0.347***
	(0.0870)	(0.0538)	(0.0740)	(0.132)
Own contribution	0.261***	0.319***	0.275***	0.739***
	(0.0416)	(0.0364)	(0.0514)	(0.0514)
Constant	-0.0269	-0.0755**	-0.0369	0.0300
	(0.0335)	(0.0358)	(0.0425)	(0.0462)
Observations	162	342	360	360

Note: Blundell-Bond estimation results. Robust standard errors clustered at session level are reported in the parentheses. ***, ** and * stand for significant level 1%, 5% and 10%, respectively.

The second finding is the coefficients of "appointed" leaders' regression results in column (3) is very close to that of "unaware" leader in column (2). It means that the "appointed" leaders do not take the influence of their own contributions into account when forming their beliefs. As "unaware" leaders could not exhibit any leading-by-example behavior, this result suggests that "appointed" leaders do not exhibit leading-by-example behavior as well, despite that they do contribute more.

The third finding is that the influence of own contributions is particularly strong in belief formation process among "legitimate" leaders in L3. "Legitimate" leaders seem to rely much more on their own contribution when forming their beliefs about the contributions of followers. Comparing "legitimate" leaders with "appointed" leaders, this result implies that the perception of leadership legitimacy changes the way leaders form and update their beliefs. When leaders are "legitimate", they tend to move away from being "naive" (basing their belief on the followers' contributions in the previous round) and put more weight on the possible influence of their own contributions on the followers. "Legitimate" leaders tend to believe that others should be more responsive to their own contributions.

To sum up, we have found that being leaders changes how people form their beliefs about contributions of followers and how they make their contribution decisions. In contribution decisions, when knowing they contribute first and their contribution would be observed, leaders act more like conditional cooperators, largely matching their contribution to their beliefs about others' contributions. "Legitimate" leaders seem to behave more conditionally cooperative than "appointed" leaders. In belief formation, while all leaders are responsive to the contribution of followers in the previous round, we also find evidence of "projection bias". "Unaware" leaders act like players in simultaneous PPG, as expected. Aware of being leaders alone doesn't make leaders believe they could influence others with their contribution, but the sense of leadership legitimacy does. "Legitimate" leaders tend to believe their own contributions have large influence on followers' contributions.

5.4 Conclusion

This study examines the effects of leading-by-example on public goods provision using sequential move PGGs. In particular, this study focuses on the leaders' behavior and how being a leader and leadership legitimacy affect leaders' behavior.

Selecting leaders and especially selecting legitimate leaders often creates selection bias. As a result, it is not easy to distinguish the impact of the perception of leadership legitimacy from the selection bias related to legitimate leaders. In this study, we use a unique experimental design to create the sense of legitimacy of leaders without creating selection biases across different types of leaders. Specifically, we use an "implicit election" procedure to select all leaders and manipulate the information that leaders receive to create variation in leaders' perceptions of leadership legitimacy.

We find that having leaders that are aware of their positions ("appointed" and "legitimate" leaders) increases contribution to the public goods. Both the leaders and the followers contribute more than in simultaneous PGGs and sequential move PGGs without leader awareness. But leaders do not enjoy high payoffs from the higher contributions. The followers are the ones who harvest most of the benefits as they tend to exploit the leaders' vulnerable position as first movers and increase their contributions to a slightly smaller extent. Surprisingly, we do not find any positive effect of leadership legitimacy on leaders' contribution levels when comparing "legitimate" leader treatment with "appointed" leader treatment.

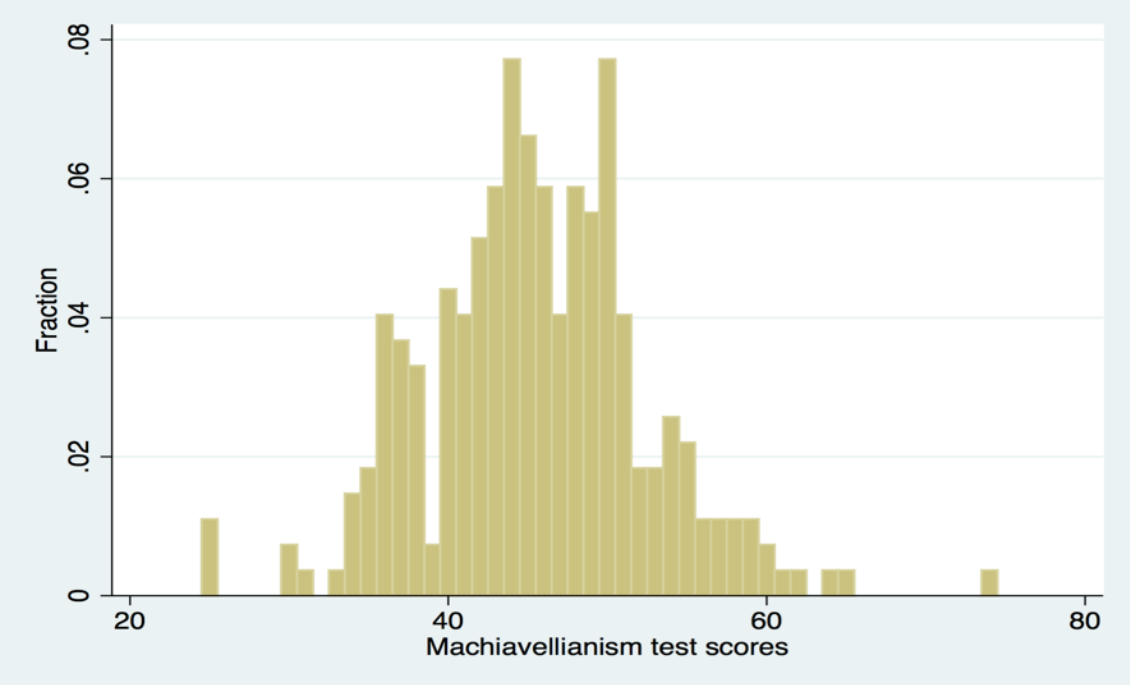
A closer look at the belief formation and contribution updating process in different treatments reveals that although leadership legitimacy does not change the contribution level of the leaders, it changes how the leaders make their contributions and update their beliefs about the contributions of the followers. Awareness of being a leader makes a leader act more conditional cooperative. And the perception of leadership legitimacy changes how legitimate leaders update their beliefs. "Legitimate" leaders tend to put much higher weight on their own contribution when updating their beliefs about the contributions of followers, while "unaware" and "appointed" leaders do not show such tendencies and they act similarly to the participants in simultaneous move PGGs. Taking own contributions into account in forming beliefs about followers implies that "legitimate" leaders believe their own contributions could affect followers' contribution decisions. This finding is consistent with the strategic use of own contributions to induce higher contributions from others in leading-by-example behavior. "Appointed" leaders show no such tendency. This implies that they do not actively try to influence others, or to "lead", but simply react to the position of being first movers whose behavior can be observed by others.

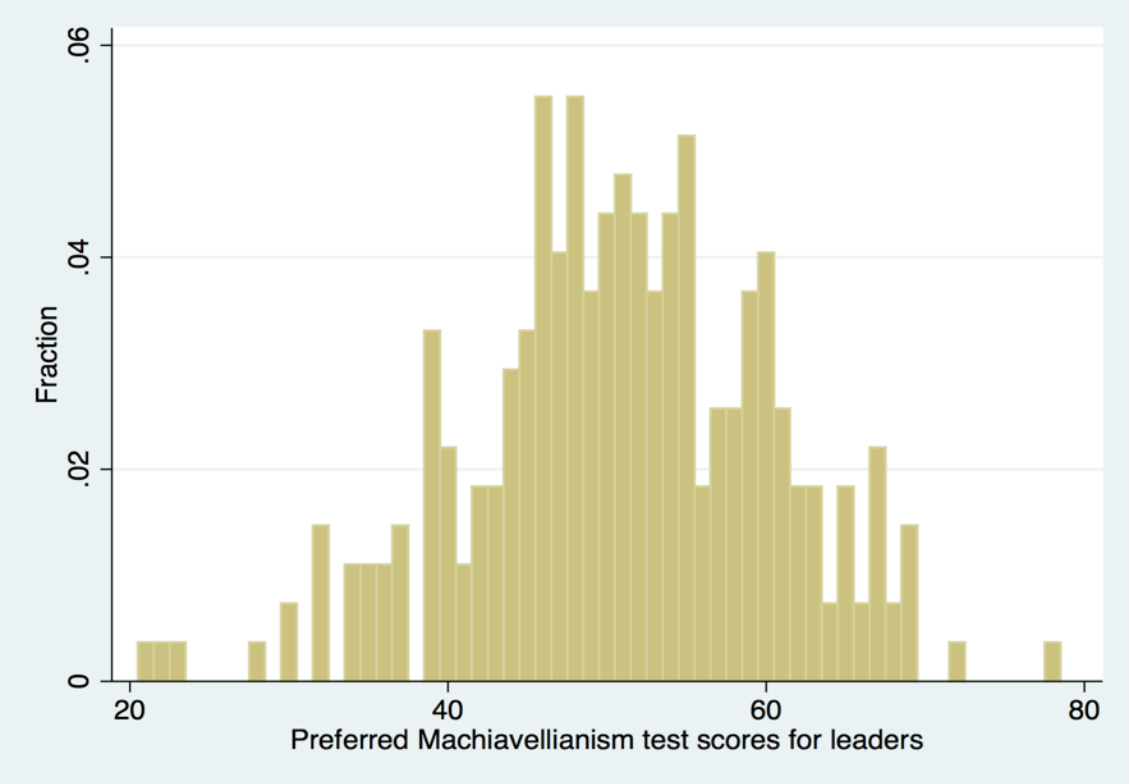
There are several limitations of this study needed to be pointed out. As our experimental design aims to overcome the self-selection issues in leader selection processes, it generates its own selection problems. Using the "implicit election" procedure, we only choose people whose Machiavellian Test scores fall in a certain range that is higher than the median level. This procedure means that the leaders we select are from a sub-group of the population. Although we observe that participants prefer more Machiavellian people as leaders in the experiment, there is no guarantee that the leader selection procedures in real life will have the same outcome regarding individual Machiavellianism. If leaders in reality have different types in terms of Machiavellianism and if people of different types react differently to leader positions and leadership legitimacy, our results may not apply. Furthermore, the self-selection of leaders in real life often carries information regarding the leaders in many other aspects, which may trigger different responses from the followers. Thus, the leader selection process itself could be the topic of interest. Future researches on the selection process of the leaders in reality are needed to provide deeper insights into the effect of leaders and leadership. Finally, in this study, we try to build the perception of leadership legitimacy though an implicit election procedure but there exist other ways to establish leadership legitimacy. People with different institutional backgrounds may react differently to leader positions and leadership legitimacy. For example, people with non-democratic backgrounds may not respond to an "election" procedure in the

same way as people from a democratic society. Thus, caution should be taken in generalizing our results to different contexts.

Appendix

A1. Distribution of Machiavellianism test scores and prefer Machiavellianism test scores for leaders





A2. English translation of the general Introduction to the experiment (originally in Chinese)

Welcome to our economics experiment at BUN Economics laboratory. Please read the instruction carefully. You understanding about the experiment will affect your final payoffs. During the experiment, please do NOT talk to others. If you have any questions, please ask the experimenter directly. If you break this rule, you will not get any payment. If you have a question, raise your hand and the experimenter will answer your question in private.

Your payoff from this experiment consists of a fixed-amount payment and the earnings from decisions in the experiment. The fixed-amount payment is 15 yuan per person. You will receive this payment regardless of your decisions in the experiment. The earnings from the experiment are determined by your decisions and the decisions of other participants in the public goods games. We will explain the concept of a public good game later. You will make in total thirty-one round of decisions, and earnings from all the decisions will add up to your final earnings. Your final payoff will be paid after the experiment in the form of cash or a Wechat/Alipay/bank transfer. The specific form of payment is your choice.

In the experiment, we use experimental tokens instead of currency unit as counting units. Your payoff will be calculated and displayed in the form of experimental tokens. After the experiment, the tokens that you earn will be exchanged into Chinese yuan. The exact exchange rate will be explained later in this introduction and will be shown on the experiment interface.

This experiment has six stages.

Stage One: Tests for individual characteristics and preferences

Stage Two: Strategy form public goods game

Stage Three: Repeated public goods game 1

Stage Four: Repeated public goods game 2

Stage Five: Repeated public goods game 3

Stage Six: Questionnaire

This experiment will take about 70 mins.

The Public Goods Game

The public goods game is the core of this experiment. Parts 2-5 of this experiment are all based on the public goods game.

In a public goods game, all participants will be randomly divided into groups of four by the computer. No one knows about the composition of the groups.

In a four-person group, every group member will receive an initial endowment of 20 tokens and she or he will determine how to allocate these tokens. You can keep all the 20 tokens in your own **private account** or contribute a part of or all the 20 tokens to a **public project**. If you decide to contribute X tokens out of the 20 tokens to the public project, the remaining 20-X tokens will be automatically kept in your private account. Your earnings from the public goods game equal the amount you gain from your private account plus the gain from the public project.

Gain from the private account

The amount kept in your private account will be your gain from the private account. For example, if you keep all 20 tokens in your private account, your gain from the private account will be 20 tokens; if you keep 6 tokens in your private account, your gain from the private account will be 6 tokens. No one else will benefit from your private account.

Gain from the public project

All members in your group will benefit from your contribution to the public project, and you will also benefit from other group members' contributions to the public project. The gain of each individual group member from the public project equals the total group contribution to the public project*0.4.

For example, if the total contribution to the public project from your group is 60 tokens, then every member of your group will gain 60*0.4=24 tokens from the public project. If the total contribution to the public project from your group is 10 tokens, then every member of your group will gain 10*0.4=4 tokens from the public project. If you do not contribute to the public project while other three group members contribute 40 tokens in total, you can still gain 40*0.4=16 tokens from the public project like the others do. Similarly, if you contribute 20 tokens to the public project while others do not contribute, they can still gain 20*0.4=8 tokens from the public project like you do.

Your total gain from the public goods game can be shown as follows: Total gained tokens =initial tokens – your contribution to the public project + 0.4*total contributions of the group to the public project.

In every round of the public goods game, besides deciding how many tokens you would like to contribute to the public project, you also need to estimate the average contributions of other group members. You can earn additional tokens based on the accuracy of your estimations. If your estimate is exactly the same as the roundup average actual contribution of others, you will gain 3 more additional tokens; if your estimate is different from the actual amount by 1 token, you will gain 2 additional tokens; if your estimate is different from the actual amount by 2 tokens, you will gain 1 additional tokens; but if your estimate is different from the actual amount by 3 or more tokens, there will be no additional earning.

After making your contribution decision and estimating others' contribution, you can see your payoffs from the public goods game before entering the next round.

Stage Two of this experiment is a strategy form public goods game. It is a one-shot game. The payoff calculation is the same as described above, but you have to make contribution decisions in two different scenarios. First, you should make you contribution decision without information about others' contributions, we call this "unconditional contribution", and estimate the average contribution of others. Second, based on every possible level of the average contribution of other group members, you need to decide your contribution strategy. For example, you need to decide how many tokens you would like to contribute to the public project if the average contribution of the other three group members is 0 token; how many tokens you would like to contribute if the average contribution of others is 1 token; how many tokens you would like to contribute if the average contribution of others is 2 tokens, and so on. When calculating payoffs, three group members will be randomly chosen and their unconditional contribution will be used as their final contributions and the last member's contribution strategy will be determined by her or his contribution strategy. In this part of the experiment, one token equals fifty cents in payoffs.

Stage Three, Stage Four and Stage Five are three repeated public good games. In each stage, you will play the public goods game for 10 rounds. At the beginning of each stage, the computer will reshuffle the group members. During the 10-round public goods game in each stage, the group composition stays the same. This means that within a stage, you will play with

the same three other participants for 10 rounds but the group members are different in different stages. In these three stages, one token equals five cents in payoffs.

The public goods games in Stage Three, Stage Four and Stage Five can be differently or be exactly the same. It depends on which experimental session you are in. Specific game instructions will be shown on the computer screen in each stage of the experiment. Please read the instructions in each stage carefully and make your decision carefully.

Practice questions

After you finish reading the general introduction of the experiment, please answer the following practice questions. These questions help you the understand how the payoffs are calculated in the public goods games and how your payoffs vary with your decisions.

1. Assume each group member has 20 tokens as initial endowment. If all group members, including you, do not contribute to the public project,

a) what is your payoff?

b) what is the payoff of other group members?

2. Assume each group member has 20 tokens as initial endowment. If you contribute 20 tokens to the public project, and the other three group members also contribute 20 tokens each,

a) what is your payoff?

b) what is the payoff of other group members?

3. Assume each group member has 20 tokens as initial endowment. If the total contribution of the other three group member is 30 tokens,

a) and you contribute 0 token, what is your payoff?

b) and you contribute 8 tokens, what is your payoff?

c) and you contribute 15 tokens, what is your payoff?

4. Assume each group member has 20 tokens as initial endowment. If you contribute 8 tokens to the public project,

a) and the total contribution of the other group members is 7 tokens, what is your payoff?

b) and the total contribution of the other group members is 12 tokens, what is your payoff?

c) and the total contribution of the other group members is 22 tokens, what is your payoff?

Chapter 6 Concluding remarks

6.1 General discussion

Cooperation is vital for solving many of the social dilemma situations in daily life, ranging from minor issues such as team assignment at school to fundamental issues such as global warming. Understanding factors affecting cooperation and developing ways to improve cooperation is of academic and societal importance. Based on the exiting literature, this thesis sets out to further study this topic from a few less studied angles and thereby contribute to the economic literature as well as to policy making by providing relevant behavioral insights.

The study starts with an assessment of the effectiveness of the eco-certification system for food crops in reducing agrochemical consumption among smallholders in China (Chapter 2). The certification system resembles a social dilemma situation, where individual farmers have incentives to free-ride on others by using more agrochemicals due to the price premium for certified products and the difficulty in monitoring farmers daily practices. Thus, the success of certification heavily relies on the voluntary compliance/cooperation from the farmers. Using panel household survey data for 4,830 households in six provinces, covering the period 2005 – 2013, we found that obtaining certificates did not reduce agrochemical consumption of farmers. Monetary incentives for free-riding can partly explain this failure.

With this case of failed cooperation among farmers in mind, we explore what affect cooperation and how to improve cooperation in the rest of the thesis. Among various factors and mechanisms that may affect cooperation, we examined three factors in particular: resource scarcity (Chapter 3), punishment and reward institutions (Chapter 4) and leading-by-example (Chapter 5). Although each of these chapters tells an independent story and examines different mechanisms, they all revolve around the role of institutions. While we directly deal with institutions in Chapter 4, we also stress the role of local irrigation management institutions in shaping cooperative norms in Chapter 3 and the sequential move natures in the leading-by-example scenario which can also be regarded as a form of institutional arrangement. Institutions are pivotal in sustaining and promoting cooperation. They do not only directly affect voluntary cooperation behavior, but also set the ground for other forces and influence cooperation behavior indirectly. It is important for researchers and policy makers to always take institutions into consideration when conducting researches and designing policies.

6.2 Cooperation and sustainable development

In Chapter 2, we evaluate the effect of adopting certified food production on agro-chemical consumption of Chinese farmers. If the certification system, which requires lower agrochemical use intensity, works well, we should expect obtaining certification to have a negative impact on agrochemical use. Yet, the challenge lies in the way certified food production is organized. Certified food production often involves a large number of farmers due to the combination of small farm size, land fragmentation and the high cost of obtaining certification. Therefore, it is difficult for the certification holders (usually farmers' cooperatives or agro-companies) and certification agencies to monitor daily agricultural activities of individual farmers in the field. Each farmer has the opportunity to deviate from the required behavior for his or her own benefits, and hence, or in other words, has the incentives to free ride on others. Therefore, the success of certified food production depends on the voluntary compliance or cooperation of the farmers. We found that adopting food certification did not significantly reduce agrochemical use. Among other factors, monetary incentives could help explain the ineffectiveness of certification. Agrochemical use increased with the relative importance of certified food production to farmers. This result implies that farmers tend to defect or free ride when having the opportunities, which would hurt the whole group and the certification system in the long run.

Our findings in Chapter 2 carry important policy implications. Combined with evidence from other case studies (e.g. Li, Zhu, and Ma 2007, Zhang 2012, Zhou and Xu 2008), Chapter 2 reveals flaws in the current agricultural product certification system in China. Despite producing under certification, farmers are not fully informed and engaged in the certification process. This contributes to the overuse of agrochemicals and the failure of certified products to reduce environmental pollution in the long term. From a governmental perspective, proper regulations on the certification procedure and strict routine inspections after issuing the certification should be established and enforced to send appropriate signals and incentives to the producers. From a producer/farmer perspective, certification holders (usually agro-companies or farmers' cooperatives) may need to design proper institutional arrangements and incentive schemes to induce cooperative behavior from farmers to overcome the temptation of free-riding.

According to our best knowledge, Chapter 2 is the first attempt to evaluate the food certification system in China on a large scale. However, despite our effort to collect data based on a national wide survey, we only have seven villages with food certification in our sample.

The small number of certified villages limits our ability to draw reliable conclusions about heterogeneous effects of different types of certification. We also do not have enough information to further explore the heterogeneous effects of different organizational structures, management styles and institutional arrangements in certified food production. What we found in Chapter 2 is the empirical evidence of something that has gone wrong. But what may be more important to the certification system, to the farmers, to the consumer and to society is how to make things right. Further studies should preferably design general principles based on successful cases and explore how to transplant these successful experiences into the certification system and other social dilemma scenarios. Chapter 3, 4 and 5 in this thesis present our efforts towards this direction, where we examine how different factors affect cooperation behavior and attempt to provide new behavioral insights into overcoming social dilemma situations.

6.3 Resource scarcity and cooperation

In Chapter 3, we examine how resource scarcity shaped the norm of cooperation in the context of irrigation agriculture. We find that long-term exposure to resource scarcity not only forces farmers to cooperate more in agricultural and irrigation activities, but also fosters a stronger norm of cooperation that spillover to other situations, i.e. contributions in a public good game. The results are consistent with the evolutionary theories of social norms that highlight the role of contextual factors. Chapter 3 contributes to a recent but emerging strand of literature about the long-term impact of the contextual factors on preferences and norms. Despite preferences and norms are usually stable in the short run, they are shaped by the environment where people interact with each other and engage in social activities.

From a policy perspective, findings in Chapter 3 underscore the importance of institutions. Irrigation is not only an agricultural technology, but also a set of institutional arrangements that demand large scale collective actions from users in order to benefit from it. An irrigation system could foster cooperation as it strengthens the interdependence among members within the community. Thus, other technological or institutional arrangements with similar features should be expected to help build cooperation in a similar way. This finding is particularly pertinent in communities facing social dilemma situations such as common pool resource management projects. The crucial element of the success of these projects is farmers' voluntary participation and contribution, which is affected by local norms or culture of cooperation. If a

project could include elements that enhance the experience or the perception of interdependence, there may be a better outcome in the long term because the project could create additional benefits from the more cooperative culture that it helps to build.

The difficulty to study the effect of resources scarcity is to find exogenous variations in resources. In reality, the current state of resources is often the outcome of past activities, which create the endogeneity issues. We take advantage of a historically formed water quota system in western China to measure differences in water scarcity in reality. While the water quota system allows us to measure exogenous variations in water scarcity, it also limits the scope of this study. It focuses on a specific type of resource (irrigation water) under a specific technology (gravity irrigation) in a specific region (a semi-arid area with moderate scarcity in irrigation water). Future studies should go beyond the context of surface water irrigation and examine whether the effects are consistent for other type of resources, other technologies and more extreme scarcity. Moreover, the quota system rules out the typical dynamics between resource stocks and resource flow/use in common pool resource management literature. Further researches on the role of scarcity in a more dynamic setting could provide more insights and implications for the common resource management issues. Finally, in Chapter 4, we are only able to test the impact of resource on in-group cooperation, but given the existence of parochial altruism (Abbink et al. 2012), the implications for out-group are not clear, especially in situation where inter-group competitions for resources are possible. More comprehensive studies that take both in-group and out-group cooperation/competition into account could provide a more complete picture of the role that resource scarcity plays in shaping cooperation.

6.4 Leading by example, leadership legitimacy and cooperation

In Chapter 5, we study the role of leading-by-example and leadership legitimacy in promoting cooperation in a sequential move public goods game. We develop a special experimental design in order to both overcome the selection bias in leader selection process and manipulate leaders' perception of the legitimacy of their leader positions. We find that having a leader improves cooperation in the public goods game. The increase of group contributions is induced by the leader's contributions. Being a leader makes leaders increase their contributions, and followers reciprocate but to a slightly smaller extent and thus they harvest the gains from increased group contributions. The perception of leadership legitimacy does not have additional impacts on the leaders' contribution level and the group contribution

level. However, we find evidence suggesting that leadership legitimacy influences how leaders make their contributions decisions and update their beliefs in the repeated public goods game. It seems that only "legitimate" leaders show the strategic use of "leading-by-example", not "appointed" leaders, whose higher contributions as leaders are simply a reaction to the sequential move game structure where their behavior can be observed by others.

The importance of leader and leadership has been stressed by literature in many fields. By acting as a role model, leaders can shape expectations of their constituency and induce them to engage in behaviors that are beneficial to the group as a whole. Chapter 5 also stresses the importance of having a leader in promoting cooperation. From a policy perspective, in order to increase private provision of public goods, it could be beneficial to mobilize a small number of participant first to set an example for other to follow, like having some seed money in charity fundraising. Moreover, results in Chapter 5 suggests that the legitimacy of the leader positions *per se* may not be as important as having leaders if the goal is to promote overall cooperation level. This result also implies that in certain scenarios, what really matters is the institutional structure {having a leader whose behavior can be observed by other}, rather than the specific way in which the institution is operated (how the leaders are selected).

While looking at the effect of leading-by-example in a sequential public good game framework like many other studies, Chapter 5 pays a special attention to leaders' decision-making process. Most studies on leading-by-example do not distinguish the effect of leading by example from the effect of game structure itself, i.e. being a first-mover, and from selection bias of leaders. While we try to address these issues with a unique experimental design in a conventional lab experiment, there are several aspects to further improve on. Compared with contribution level of Chinese farmers (see Chapter 3 and Chapter 4), Chinese college students tend to be less cooperative to begin with and more responsive in the game as their contributions show sharper decline trend over time. This difference suggests that we should be cautious if we attempt to generalize our findings in the lab to the field. Similarly, although we observe that participants prefer more Machiavellian people as leaders in the experiment, there is no guarantee that the leader selection procedures in real life will have the same outcome regarding individual Machiavellianism. Future studies in the field are necessary if we want to draw solid conclusions to guide policies in the field. Moreover, as we try to control the selection process of the leaders in the lab to isolate the effect of leadership legitimacy itself, the self-selection of leaders in reality is almost always intertwined with the leadership legitimacy and bares information about the leaders. Thus, the self-selection process *per se* could be the topic of

interest. Future researches on the selection process of the leaders in reality can provide deeper insights into the effect of leaders and leadership.

6.5 Institutions and cooperation

Institutions shape people's cooperation behavior through both the informal constrains (e.g. social norms) and formal rules (e.g. regulations and laws). While findings in both Chapter 3 and Chapter 5 indirectly feature the role of institutions in shaping cooperative norms and promote cooperative behavior, we deal with the formal institutions directly in Chapter 4. Specifically, we examine how the interaction between preferences and implemented institutions affect cooperation using a lab-in-the-field experiment with Chinese farmers as subjects. We find that while subjects have different institutional preferences, their institutional preferences do not significantly affect the effectiveness of implemented institutions. Moreover, it is puzzling to find that the preference for the punishment institution is positively correlated with being free-riders. The link between certain "anti-social" or "efficiency-reducing" social preference profiles and the preference for the punishment institution may provide an explanation for the puzzling observation.

The lack of response to the imposed institutions may the results of the typical top-down governing regimes in rural China. Farmers who are used to having polices imposed on them could have learnt not to take their own preferences into consideration. It implies that, in certain context, exogenously policies or policy changes may have the same effect regardless peoples' preferences. Moreover, the preferences for certain institutions may provide some insights into endogenous policy-making process. Active support of a particular institution or policy from certain groups in the society does not necessarily represent their commitment to socially desirable outcomes that the institution or policy in question aims to achieve, and their willingness to comply if it is ever implemented. The preference for an institution could be driven by certain underlying preferences, beliefs or ideologies, instead of the beliefs in the effectiveness of such institution in achieving the goals.

While Chapter 4 offers insights into the relationship between institutions, preferences and cooperation, we again would like to be cautious when generalizing the results to other situations. The institutions in the experiment are a punishment institution and a reward institution, and the punishment and reward institutions are designed to be formal (from the experimenter, not other subjects), non-deterrent (the incentive of free-riding) and exogenously

imposed (regardless their institutional preferences). The lack of response to the institution mismatch might not apply to other institutional settings. Future studies should expand to alternatives institutions and examine whether the relationship between institutional preferences and the effect of implemented institutions still hold, especially for endogenously determined institutions. Moreover, when using farmers as experimental subjects, we are looking at farmers carrying experiences in their own context. The context of irrigation agriculture may contribute to non-impact findings, because irrigation is found be associated with concentration of power and less democratic attitudes (Bentzen, Kaarsen, and Wingender 2017). People from other background could react differently, which is left for studies in the future. Finally, the relationship between preferences for the punishment institution and "anti-social" or "efficiency-reducing" social preference profiles is only correlational, not causal. Future studies are needed to uncover the mechanisms behind this correlation.

6.6 Obstacles of cooperation

A large part of this thesis is devoted to explore and examine the factors and mechanisms that can improve and sustain voluntary cooperation among individuals facing free-riding incentives. While this thesis has shown how resource scarcity, institutional preferences and leading-by-example are associated with voluntary cooperation behavior, as the other side of the coin, it also reveals various obstacles that undermines cooperation.

Self-interested incentives are still driving the free-riding behavior, which can be found both in the field (Chapter 2) and in the lab (Chapter 4 and 5). Moreover, findings in Chapter 3 imply that an environment that reduces the interdependence among individuals, such as resource abundance, could undermine the norm of cooperation in the long run. The existence of people with "efficiency-reducing" or "anti-social" social preference profiles is another obstacle for cooperation among individuals (Chapter 4). People with such social preference features do not even response well to monetary incentives for themselves. Thus, how to motivate or discipline these sub-group of people is important in improving and sustaining cooperation. While having a leader increases cooperation, followers do not fully reciprocate to leaders' higher contributions. Followers tend to increase their contributions to a less extend so that they can harvest most of the benefits gained from the increased total contributions to the public good, while leaders do not better off despite their vital roles in promoting group contributions (Chapter 5). This phenomenon is referred as "leader's curse". Without proper

rewards for their behavior, leaders could be discouraged and reduce their contributions later and the cooperation would be gradually eroded as we have seen in Chapter 5. Therefore, the incomplete reciprocity or incomplete conditional cooperation represents another mechanism that undermines a potentially effective way to promote voluntary cooperation.

Knowledge about these obstacles can help us to understand the potential difficulties in improving voluntary cooperation. Contexts, institutions and preferences all influence individual's cooperation behavior. The effectiveness of one particular instrument may vary with these factors. Policies that relies on voluntary cooperation and policies instruments that aim to induce voluntary cooperation should take these factors into consideration.

6.6 Final remarks

This thesis starts with an attempt to assess the effectiveness of eco-certification of agricultural products in reducing agro-chemical use. Cooperation among farmers is vital in the success of such initiatives and yet we find that the certification system does work well and monetary incentives of free riding is one of the reasons why the certification system has not been very successful. Then, we investigate how different contexts, institutions and preferences affect cooperation combining data from a lab experiment, a lab-in-the-field experiment and a rural household survey. While this thesis has examined what affect cooperation and how to improve cooperation when facing free-riding incentives and provided new insights, it only taps in the surface of this topic and raises many news questions for future researches. One particular interesting question is how to assess the short-term and the long-term effects of a certain policy instrument in promoting cooperation. While short term interventions, e.g. introducing a new institutional arrangement, can induce behavior changes, they may also have long term implications regarding changes in social norms and preferences which in return may enhance or offset the short-term effects. While studies in this thesis are conducted in various contexts in China, we believe that the findings discover some general relationships that also apply to similar contexts in other countries. I would like to test them in future studies. Yet, as this thesis stresses the importance of the contextual factors, we are cautious to claim that the exact findings can be expected from any other contexts. A successful polices to promote voluntary cooperation should be customized according to local contextual environments, existing institutions and cultural backgrounds.

References

Abbink, K., J. Brandts, B. Herrmann, and H. Orzen. 2012. "Parochial altruism in inter-group conflicts." *Economics Letters* 117 (1):45-48. doi: 10.1016/j.econlet.2012.04.083.

Abbink, K., and A. Sadrieh. 2009. "The pleasure of being nasty." *Economics Letters* 105 (3):306-308. doi: 10.1016/j.econlet.2009.08.024.

Aflaki, S. 2013. "The effect of environmental uncertainty on the tragedy of the commons." *Games and Economic Behavior* 82:240-253. doi: 10.1016/j.geb.2013.07.011.

Alesina, A., P. Giuliano, and N. Nunn. 2013. "On the Origins of Gender Roles: Women and the Plough." *Quarterly Journal of Economics* 128 (2):469-530. doi: 10.1093/qje/qjt005.

Andreoni, J., and L. K. Gee. 2012. "Gun for hire: Delegated enforcement and peer punishment in public goods provision." *Journal of Public Economics* 96 (11-12):1036-1046. doi: 10.1016/j.jpubeco.2012.08.003.

Aoyagi, Keitaro, Yasuyuki Sawada, and Masahiro Shoji. 2014. Does Infrastructure Facilitate Social Capital Accumulation? Evidence from Natural and Artefactual Field Experiments in a Developing Country. JICA Research Institute.

Araral, E. 2009. "What Explains Collective Action in the Commons? Theory and Evidence from the Philippines." *World Development* 37 (3):687-697.

Attanasio, O., S. Polania-Reyes, and L. Pellerano. 2015. "Building social capital: Conditional cash transfers and cooperation." *Journal of Economic Behavior & Organization* 118:22-39. doi: 10.1016/j.jebo.2015.04.004.

Bardhan, P. 1993. "Analytics of the Institutions of Informal Cooperation in Rural-Development." *World Development* 21 (4):633-639. doi: Doi 10.1016/0305-750x(93)90115-P.

Bardhan, P. 2000. "Irrigation and cooperation: An empirical analysis of 48 irrigation communities in South India." *Economic Development and Cultural Change* 48 (4):847-865. doi: Doi 10.1086/452480.

Benjamin, D., L. Brandt, and J. Giles. 2005. "The evolution of income inequality in rural China." *Economic Development and Cultural Change* 53 (4):769-824.

Bentzen, J. S., N. Kaarsen, and A. M. Wingender. 2017. "Irrigation and Autocracy." *Journal of the European Economic Association* 15 (1):1-53. doi: 10.1111/jeea.12173.

Bernhard, H., U. Fischbacher, and E. Fehr. 2006. "Parochial altruism in humans." *Nature* 442 (7105):912-915.

Beuchelt, T. D., and M. Zeller. 2011. "Profits and poverty: Certification's troubled link for Nicaragua's organic and fairtrade coffee producers." *Ecological Economics* 70 (7):1316-1324.

Blackman, A., and M. A. Naranjo. 2012. "Does eco-certification have environmental benefits? Organic coffee in Costa Rica." *Ecological Economics* 83:58-66.

Blanco, E., M. C. Lopez, and S. Villamayor-Tomas. 2015. "Exogenous degradation in the commons: Field experimental evidence." *Ecological Economics* 120:430-439. doi: 10.1016/j.ecolecon.2015.03.028.

Blundell, R., and S. Bonds. 1998. "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics* 87 (1):115-143. doi: Doi 10.1016/S0304-4076(98)00009-8.

Botelho, A., A. Dinar, L. M. C. Pinto, and A. Rapoport. 2015. "Promoting cooperation in resource dilemmas: Theoretical predictions and experimental evidence." *Journal of Behavioral and Experimental Economics* 54:40-49. doi: 10.1016/j.socec.2014.11.007.

Botelho, Anabela, Glenn W Harrison, Lígia Pinto, and Elisabet E Rutström. 2005. "Social norms and social choice." *Working Paper*.

Bowles, S. 1998. "Endogenous preferences: The cultural consequences of markets and other economic institutions." *Journal of Economic Literature* 36 (1):75-111.

Boyd, Robert, and Peter J. Richerson. 2005. *The origin and evolution of cultures, Evolution and cognition*. Oxford ; New York: Oxford University Press.

Brandts, J., D. J. Cooper, and R. A. Weber. 2015. "Legitimacy, Communication, and Leadership in the Turnaround Game." *Management Science* 61 (11):2627-2645. doi: 10.1287/mnsc.2014.2021.

Brooks, J. S. 2010. "The Buddha mushroom: Conservation behavior and the development of institutions in Bhutan." *Ecological Economics* 69 (4):779-795. doi: 10.1016/j.ecolecon.2008.01.022.

Brosig-Koch, J., C. Helbach, A. Ockenfels, and J. Weimann. 2011. "Still different after all these years: Solidarity behavior in East and West Germany." *Journal of Public Economics* 95 (11-12):1373-1376. doi: 10.1016/j.jpubeco.2011.06.002.

Brunnschweiler, C. N., and E. H. Bulte. 2009. "Natural resources and violent conflict: resource abundance, dependence, and the onset of civil wars." *Oxford Economic Papers-New Series* 61 (4):651-674. doi: 10.1093/oep/gpp024.

Bulte, E., Q. Tu, and J. List. 2015. "Battle of the Sexes: How Sex Ratios Affect Female Bargaining Power." *Economic Development and Cultural Change* 64 (1):143-161. doi: 10.1086/682706.

Burke, M., S. M. Hsiang, and E. Miguel. 2015. "Climate and Conflict." In *Annual Review of Economics, Vol 7*, edited by K. J. Arrow and T. F. Bresnahan, 577-+.

Carpenter, Jeffrey, and Erika Seki. 2011. "Do Social Preferences Increase Productivity? Field Experimental Evidence from Fishermen in Toyama Bay." *Economic Inquiry* 49 (2):612-630. doi: 10.1111/j.1465-7295.2009.00268.x.

Cassar, A., A. Healy, and C. Von Kessler. 2017. "Trust, Risk, and Time Preferences After a Natural Disaster: Experimental Evidence from Thailand." *World Development* 94:90-105.

Cecchi, F., and J. Duchoslav. 2018. "The effect of prenatal stress on cooperation: Evidence from violent conflict in Uganda." *European Economic Review* 101:35-56. doi: 10.1016/j.euroecorev.2017.09.015.

Centre, China Green Food Development. 2015. Annual Green Food Statistics Report 2014. edited by China Green Food Development Centre. Beijing.

Chattopadhyay, R., and E. Duflo. 2004. "Women as policy makers: Evidence from a randomized policy experiment in India." *Econometrica* 72 (5):1409-1443. doi: DOI 10.1111/j.1468-0262.2004.00539.x.

Chen, Y., and S. X. Li. 2009. "Group Identity and Social Preferences." *American Economic Review* 99 (1):431-457. doi: 10.1257/aer.99.1.431.

Cheng, C. W., Y. Shi, and T. J. Wen. 2010. Beijing.

Choi, J. K., and T. K. Ahn. 2013. "Strategic reward and altruistic punishment support cooperation in a public goods game experiment." *Journal of Economic Psychology* 35:17-30. doi: 10.1016/j.joep.2013.01.001.

Christie, Richard, and Florence L Geis. 1970. *Studies in machiavellianism*: Academic Press.

Cinyabuguma, M., T. Page, and L. Putterman. 2006. "Can second-order punishment deter perverse punishment?" *Experimental Economics* 9 (3):265-279.

Dal Bo, P., A. Foster, and L. Putterman. 2010. "Institutions and Behavior: Experimental Evidence on the Effects of Democracy." *American Economic Review* 100 (5):2205-2229.

Dannenberg, A. 2015. "Leading by example versus leading by words in voluntary contribution experiments." *Social Choice and Welfare* 44 (1):71-85. doi: 10.1007/s00355-014-0817-8.

Dickinson, D. L., E. G. Dutcher, and C. S. Rodet. 2015. "Observed punishment spillover effects: a laboratory investigation of behavior in a social dilemma." *Experimental Economics* 18 (1):136-153. doi: 10.1007/s10683-014-9399-7.

Dickinson, D. L., D. Masclet, and M. C. Villeval. 2015. "Norm enforcement in social dilemmas: An experiment with police commissioners." *Journal of Public Economics* 126:74-85. doi: 10.1016/j.jpubeco.2015.03.012.

Drouvelis, Michalis, and Julian C. Jamison. 2015. "Selecting public goods institutions: Who likes to punish and reward?" *Southern Economic Journal* 82 (2):501-534. doi: 10.1002/soej.12049.

Duchoslav, J. 2017. "Prenatal Temperature Shocks Reduce Cooperation: Evidence from Public Goods Games in Uganda." *Frontiers in Behavioral Neuroscience* 11. doi: 10.3389/fnbeh.2017.00249.

Ertan, A., T. Page, and L. Putterman. 2009. "Who to punish? Individual decisions and majority rule in mitigating the free rider problem." *European Economic Review* 53 (5):495-511.

Falk, A., and J. J. Heckman. 2009. "Lab experiments are a major source of knowledge in the social sciences." *Science* 326 (5952):535-8. doi: 10.1126/science.1168244.

Falk, A., S. Meier, and C. Zehnder. 2013. "Do Lab Experiments Misrepresent Social Preferences? The Case of Self-Selected Student Samples." *Journal of the European Economic Association* 11 (4):839-852.

Fehr, E., and S. Gachter. 2000. "Cooperation and punishment in public goods experiments." *American Economic Review* 90 (4):980-994. doi: DOI 10.1257/aer.90.4.980.

Fehr, E., and S. Gachter. 2002. "Altruistic punishment in humans." *Nature* 415 (6868):137-40. doi: 10.1038/415137a.

Figuieres, C., D. Masclet, and M. Willinger. 2012. "Vanishing Leadership and Declining Reciprocity in a Sequential Contribution Experiment." *Economic Inquiry* 50 (3):567-584. doi: 10.1111/j.1465-7295.2011.00415.x.

Fischbacher, U. 2007. "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental Economics* 10 (2):171-178. doi: 10.1007/s10683-006-9159-4.

Fischbacher, U., and S. Gachter. 2010. "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments." *American Economic Review* 100 (1):541-556. doi: 10.1257/aer.100.1.541.

Frey, B. S., and S. Meier. 2004. "Social comparisons and pro-social behavior: Testing "conditional cooperation" in a field experiment." *American Economic Review* 94 (5):1717-1722. doi: Doi 10.1257/0002828043052187.

Fujiie, M., Y. Hayami, and M. Kikuchi. 2005. "The conditions of collective action for local commons management: the case of irrigation in the Philippines." *Agricultural Economics* 33 (2):179-189. doi: 10.1111/j.1574-0862.2005.00351.x.

Gachter, S., D. Nosenzo, E. Renner, and M. Sefton. 2010. "Sequential vs. simultaneous contributions to public goods: Experimental evidence." *Journal of Public Economics* 94 (7-8):515-522. doi: 10.1016/j.jpubeco.2010.03.002.

Gachter, S., and E. Renner. 2010. "The effects of (incentivized) belief elicitation in public goods experiments." *Experimental Economics* 13 (3):364-377. doi: 10.1007/s10683-010-9246-4.

Gächter, Simon, and Elke Renner. 2014. "Leaders as role models for the voluntary provision of public goods." *Working Paper*.

Gambelli, D., F. Solfanelli, R. Zanoli, A. Zorn, C. Lippert, and S. Dabbert. 2014. "Non-compliance in organic farming: A cross-country comparison of Italy and Germany." *Food Policy* 49:449-458.

Gatiso, T. T., B. Vollan, and E. A. Nuppenau. 2015. "Resource scarcity, and democratic elections in commons dilemmas: An experiment on forest use in Ethiopia." *Ecological Economics* 114:199-207. doi: 10.1016/j.ecolecon.2015.04.005.

Gneezy, U., A. Leibbrandt, and J. A. List. 2016. "Ode to the Sea: Workplace Organizations and Norms of Cooperation." *Economic Journal* 126 (595):1856-1883. doi: 10.1111/ecoj.12209.

Guo, J. H., X. J. Liu, Y. Zhang, J. L. Shen, W. X. Han, W. F. Zhang, P. Christie, K. W. T. Goulding, P. M. Vitousek, and F. S. Zhang. 2010. "Significant Acidification in Major Chinese Croplands." *Science* 327 (5968):1008-1010. doi: 10.1126/science.1182570.

Gurerk, O. 2013. "Social learning increases the acceptance and the efficiency of punishment institutions in social dilemmas." *Journal of Economic Psychology* 34:229-239. doi: 10.1016/j.joep.2012.10.004.

Gurerk, O., B. Irlenbusch, and B. Rockenbach. 2006. "The competitive advantage of sanctioning institutions." *Science* 312 (5770):108-11. doi: 10.1126/science.1123633.

Guth, W., M. V. Levati, M. Sutter, and E. Van der Heijden. 2007. "Leading by example with and without exclusion power in voluntary contribution experiments." *Journal of Public Economics* 91 (5-6):1023-1042. doi: 10.1016/j.jpubeco.2006.10.007.

Haigner, Stefan D., and Florian Wakolbinger. 2010. "To lead or not to lead." *Economics Letters* 108 (1):93-95. doi: 10.1016/j.econlet.2010.04.025.

Han, Y. . 2010. "The evolution of Chinese green food industry and its phase characteristics and development strategies." *Chinese Rural Economy* 2010 (2):33-43.

Harrison, G. W., and J. A. List. 2004. "Field experiments." *Journal of Economic Literature* 42 (4):1009-1055. doi: Doi 10.1257/0022051043004577.

Heineck, G., and B. Sussmuth. 2013. "A different look at Lenin's legacy: Social capital and risk taking in the Two Germanies." *Journal of Comparative Economics* 41 (3):789-803. doi: 10.1016/j.jce.2013.02.005.

Herrmann, B., C. Thoni, and S. Gachter. 2008. "Antisocial punishment across societies." *Science* 319 (5868):1362-7. doi: 10.1126/science.1153808.

Holt, Charles A. 2007. *Markets, games, & strategic behavior*. Boston: Pearson Addison Wesley.

Ito, J. 2012. "Collective Action for Local Commons Management in Rural Yunnan, China: Empirical Evidence and Hypotheses Using Evolutionary Game Theory." *Land Economics* 88 (1):181-200.

Jack, B. K., and M. P. Recalde. 2015. "Leadership and the voluntary provision of public goods: Field evidence from Bolivia." *Journal of Public Economics* 122:80-93. doi: 10.1016/j.pubeco.2014.10.003.

Jennings, C., and S. Sanchez-Pages. 2017. "Social capital, conflict and welfare." *Journal of Development Economics* 124:157-167.

Jones, B. F., and B. A. Olken. 2005. "Do leaders matter? National leadership and growth since world war II." *Quarterly Journal of Economics* 120 (3):835-864.

Kamei, Kenju, Louis Putterman, and Jean-Robert Tyran. 2014. "State or nature? Endogenous formal versus informal sanctions in the voluntary provision of public goods." *Experimental Economics* 18 (1):38-65. doi: 10.1007/s10683-014-9405-0.

Kleemann, Linda, and Awudu Abdulai. 2013. "Organic certification, agro-ecological practices and return on investment: Evidence from pineapple producers in Ghana." *Ecological Economics* 93:330-341.

Kocher, M. G., and D. Matzat. 2016. "Preferences over Punishment and Reward Mechanisms in Social Dilemmas." *Journal of Institutional and Theoretical Economics-Zeitschrift Fur Die Gesamte Staatswissenschaft* 172 (1):163-194. doi: 10.1628/093245616x14500948554072.

Kuhn, A. 2013. "Inequality Perceptions, Distributional Norms, and Redistributive Preferences in East and West Germany." *German Economic Review* 14 (4):483-499. doi: 10.1111/j.1468-0475.2012.00577.x.

Le, C., Y. Zha, Y. Li, D. Sun, H. Lu, and B. Yin. 2010. "Eutrophication of Lake Waters in China: Cost, Causes, and Control." *Environmental Management* 45 (4):662-668. doi: 10.1007/s00267-010-9440-3.

Ledyard, O. 1995. "Public goods: some experimental results." In *Handbook of experimental economics*, edited by J. Kagel and A. Roth. Princeton: Princeton University Press.

Leibbrandt, A., U. Gneezy, and J. A. List. 2013. "Rise and fall of competitiveness in individualistic and collectivistic societies." *Proc Natl Acad Sci U S A* 110 (23):9305-8. doi: 10.1073/pnas.1300431110.

Levati, M. V., M. Sutter, and E. van der Heijden. 2007. "Leading by example in a public goods experiment with heterogeneity and incomplete information." *Journal of Conflict Resolution* 51 (5):793-818. doi: 10.1177/0022002707302796.

Levitt, S. D., and J. A. List. 2007. "What do laboratory experiments measuring social preferences reveal about the real world?" *Journal of Economic Perspectives* 21 (2):153-174.

Levy, D. M., K. Padgitt, S. J. Peart, D. Houser, and E. Xiao. 2011. "Leadership, cheap talk and really cheap talk." *Journal of Economic Behavior & Organization* 77 (1):40-52. doi: 10.1016/j.jebo.2010.02.018.

Li, G. S., L. L. Zhu, and L. Ma. 2007. "The impact of Non-hazard certification on the pesticide use of rural households: the case of Nanjing, Jiangsu province." *Rural Economy* 2007 (5):95-97.

Lin, L., D. Y. Zhou, and C. X. Ma. 2010. "Green food industry in China: development, problems and policies." *Renewable Agriculture and Food Systems* 25 (1):69-80. doi: 10.1017/S174217050999024x.

Lindahl, T., A. S. Crepin, and C. Schill. 2016. "Potential Disasters can Turn the Tragedy into Success." *Environmental & Resource Economics* 65 (3):657-676.

List, J. A., and D. Lucking-Reiley. 2002. "The effects of seed money and refunds on charitable giving: Experimental evidence from a university capital campaign." *Journal of Political Economy* 110 (1):215-233. doi: Doi 10.1086/324392.

Maystadt, J. F., and O. Ecker. 2014. "Extreme Weather and Civil War: Does Drought Fuel Conflict in Somalia through Livestock Price Shocks?" *American Journal of Agricultural Economics* 96 (4):1157-1182.

Miguel, E., S. Satyanath, and E. Sergenti. 2004. "Economic shocks and civil conflict: An instrumental variables approach." *Journal of Political Economy* 112 (4):725-753. doi: Doi 10.1086/421174.

Moxnes, E., and E. van der Heijden. 2003. "The effect of leadership in a public bad experiment." *Journal of Conflict Resolution* 47 (6):773-795. doi: 10.1177/0022002703258962.

Nie, Zihan, and Xiaojun Yang. 2018. "Resource scarcity and cooperation: evidence from an irrigation district in western China." *Working Paper*.

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.

North, D. C. 1991. "Institutions." *Journal of Economic Perspectives* 5 (1):97-112. doi: DOI 10.1257/jep.5.1.97.

Nowak, M. A., and Roger Highfield. 2011. *SuperCooperators : altruism, evolution, and why we need each other to succeed*. 1st Free Press hardcover ed. New York: Free Press.

Oelofse, M., H. Hogh-Jensen, L. S. Abreu, G. F. Almeida, Q. Y. Hui, T. Sultan, and A. de Neergaard. 2010. "Certified organic agriculture in China and Brazil: Market accessibility and outcomes following adoption." *Ecological Economics* 69 (9):1785-1793.

Ortega, David L., H. Holly Wang, Laping Wu, and Soo Jeong Hong. 2015. "Retail channel and consumer demand for food quality in China." *China Economic Review* 36:359-366. doi: 10.1016/j.chieco.2015.04.005.

Oses-Eraso, N., F. Udina, and M. Viladrich-Grau. 2008. "Environmental versus human-induced scarcity in the commons: Do they trigger the same response?" *Environmental & Resource Economics* 40 (4):529-550.

Oses-Eraso, N., and M. Viladrich-Grau. 2007. "Appropriation and concern for resource scarcity in the commons: An experimental study." *Ecological Economics* 63 (2-3):435-445. doi: 10.1016/j.ecolecon.2006.11.016.

Ostrom, E. 2009. "A general framework for analyzing sustainability of social-ecological systems." *Science* 325 (5939):419-22. doi: 10.1126/science.1172133.

Ostrom, E., and R. Gardner. 1993. "Coping with Asymmetries in the Commons - Self-Governing Irrigation Systems Can Work." *Journal of Economic Perspectives* 7 (4):93-112.

Ostrom, Elinor. 1990. *Governing the commons : the evolution of institutions for collective action, The Political economy of institutions and decisions*. Cambridge ; New York: Cambridge University Press.

Pfaff, A., M. A. Velez, P. A. Ramos, and A. Molina. 2015. "Framed field experiment on resource scarcity & extraction: Path-dependent generosity within sequential water

appropriation." *Ecological Economics* 120:416-429. doi: 10.1016/j.ecolecon.2015.06.002.

Potters, J., M. Sefton, and L. Vesterlund. 2007. "Leading-by-example and signaling in voluntary contribution games: an experimental study." *Economic Theory* 33 (1):169-182. doi: 10.1007/s00199-006-0186-3.

Prediger, S., B. Vollan, and B. Herrmann. 2014. "Resource scarcity and antisocial behavior." *Journal of Public Economics* 119:1-9. doi: 10.1016/j.jpubeco.2014.07.007.

Putterman, L., J. R. Tyran, and K. Kamei. 2011. "Public goods and voting on formal sanction schemes." *Journal of Public Economics* 95 (9-10):1213-1222. doi: 10.1016/j.jpubeco.2011.05.001.

Qu, F. T., A. Kuyvenhoven, X. P. Shi, and N. Heerink. 2011. "Sustainable natural resource use in rural China: Recent trends and policies." *China Economic Review* 22 (4):444-460.

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

Reuben, Ernesto, and Krisztina Timko. 2017. "On the Effectiveness of Elected Male and Female Leaders and Team Coordination."

Rivas, M. F., and M. Sutter. 2011. "The benefits of voluntary leadership in experimental public goods games." *Economics Letters* 112 (2):176-178. doi: 10.1016/j.econlet.2011.04.007.

Ross, L., D. Greene, and P. House. 1977. "False Consensus Effect - Egocentric Bias in Social-Perception and Attribution Processes." *Journal of Experimental Social Psychology* 13 (3):279-301. doi: Doi 10.1016/0022-1031(77)90049-X.

Sanders, R. 2006. "A market road to sustainable agriculture? Ecological agriculture, green food and organic agriculture in China." *Development and Change* 37 (1):201-226.

Sarangi, Sudipta, Chandan Jha, and Gautam Hazarika. 2015. The Role of Historical Resource Scarcity in Modern Gender Inequality.

Scott, S., Z. Z. Si, T. Schumilas, and A. J. Chen. 2014. "Contradictions in state- and civil society-driven developments in China's ecological agriculture sector." *Food Policy* 45:158-166.

Sefton, M., R. Shupp, and J. M. Walker. 2007. "The effect of rewards and sanctions in provision of public goods." *Economic Inquiry* 45 (4):671-690. doi: 10.1111/j.1465-7295.2007.00051.x.

Sheng, J. P., L. Shen, Y. H. Qiao, M. M. Yu, and B. Fan. 2009. "Market trends and accreditation systems for organic food in China." *Trends in Food Science & Technology* 20 (9):396-401.

Smith, A. 2013. "Estimating the causal effect of beliefs on contributions in repeated public good games." *Experimental Economics* 16 (3):414-425. doi: 10.1007/s10683-012-9345-5.

Solstad, J. T., and K. A. Brekke. 2011. "Does the Existence of a Public Good Enhance Cooperation among Users of Common-Pool Resources?" *Land Economics* 87 (2):335-345.

Sun, B., L. X. Zhang, L. Z. Yang, F. S. Zhang, D. Norse, and Z. L. Zhu. 2012. "Agricultural Non-Point Source Pollution in China: Causes and Mitigation Measures." *Ambio* 41 (4):370-379. doi: 10.1007/s13280-012-0249-6.

Sutter, M., S. Haigner, and M. G. Kocher. 2010. "Choosing the Carrot or the Stick? Endogenous Institutional Choice in Social Dilemma Situations." *Review of Economic Studies* 77 (4):1540-1566. doi: 10.1111/j.1467-937X.2010.00608.x.

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. doi: 10.1126/science.1246850.

Thogersen, J., M. D. de Barcellos, M. G. Perin, and Y. F. Zhou. 2015. "Consumer buying motives and attitudes towards organic food in two emerging markets China and Brazil." *International Marketing Review* 32 (3-4):389-413.

Thoni, C. 2014. "Inequality aversion and antisocial punishment." *Theory and Decision* 76 (4):529-545. doi: 10.1007/s11238-013-9382-3.

Tsusaka, T. W., K. Kajisa, V. O. Pede, and K. Aoyagi. 2015. "Neighborhood effects and social behavior: The case of irrigated and rainfed farmers in Bohol, the Philippines." *Journal of Economic Behavior & Organization* 118:227-246. doi: 10.1016/j.jebo.2015.04.022.

Tversky, A., and D. Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185 (4157):1124-31. doi: 10.1126/science.185.4157.1124.

Tyran, J. R., and L. P. Feld. 2006. "Achieving compliance when legal sanctions are non-deterrent." *Scandinavian Journal of Economics* 108 (1):135-156. doi: 10.1111/j.1467-9442.2006.00444.x.

Vollan, B., A. Landmann, Y. X. Zhou, B. L. Hu, and C. Herrmann-Pillath. 2017. "Cooperation and authoritarian values: An experimental study in China." *European Economic Review* 93:90-105. doi: 10.1016/j.euroecorev.2017.01.007.

von Carnap, T. 2017. "Irrigation as a Historical Determinant of Social Capital in India? A Large-Scale Survey Analysis." *World Development* 95:316-333.

von Rueden, C., M. Gurven, H. Kaplan, and J. Stieglitz. 2014. "Leadership in an egalitarian society." *Hum Nat* 25 (4):538-66. doi: 10.1007/s12110-014-9213-4.

Walumbwa, F. O., B. J. Avolio, W. L. Gardner, T. S. Wernsing, and S. J. Peterson. 2008. "Authentic leadership: Development and validation of a theory-based measure." *Journal of Management* 34 (1):89-126. doi: 10.1177/0149206307308913.

Wang, D. Z., D. W. Zhao, and H. Y. Du. 2009. "Optimization and policy innovation of Chinese green food industrial structure." *China Industrial Economics* 2009 (9):67-76.

Wang, Y. H., C. L. Chen, and E. Araral. 2016. "The Effects of Migration on Collective Action in the Commons: Evidence from Rural China." *World Development* 88:79-93. doi: 10.1016/j.worlddev.2016.07.014.

Wu, Y. R. 2011. "Chemical fertilizer use efficiency and its determinants in China's farming sector Implications for environmental protection." *China Agricultural Economic Review* 3 (2):117-130.

Xie, H. L., and G. Y. Liu. 2015. "Spatiotemporal differences and influencing factors of multiple cropping index in China during 1998-2012." *Journal of Geographical Sciences* 25 (11):1283-1297. doi: 10.1007/s11442-015-1234-3.

Yang, X. J., and F. Carlsson. 2016. "Influence and choice shifts in households: An experimental investigation." *Journal of Economic Psychology* 53:54-66. doi: 10.1016/j.joep.2015.11.002.

Yao, Y. . 2011. Description of the NFS.

Yin, S. J., M. Chen, Y. J. Xu, and Y. S. Chen. 2017. "Chinese consumers' willingness-to-pay for safety label on tomato: evidence from choice experiments." *China Agricultural Economic Review* 9 (1):141-155.

Yin, S. J., L. H. Wu, L. Du, and M. Chen. 2010. "Consumers' purchase intention of organic food in China." *Journal of the Science of Food and Agriculture* 90 (8):1361-1367.

Yu, X. H., Z. F. Gao, and Y. C. Zeng. 2014. "Willingness to pay for the "Green Food" in China." *Food Policy* 45:80-87.

Yu, Xiaohua, and Guoqing Zhao. 2009. "Chinese agricultural development in 30 years: A literature review." *Frontiers of Economics in China* 4 (4):633-648.

Zhang, J. L., F. R. Zhang, L. Q. Zhang, and W. Wang. 2009. "Transaction Costs in Water Markets in the Heihe River Basin in Northwest China." *International Journal of Water Resources Development* 25 (1):95-105. doi: Pii 908041585

10.1080/07900620802517541.

Zhang, L., N. Heerink, L. Dries, and X. P. Shi. 2013. "Water users associations and irrigation water productivity in northern China." *Ecological Economics* 95:128-136. doi: 10.1016/j.ecolecon.2013.08.014.

Zhang, T. . 2012. "An analysis on the influencing factor for the farming households' green vegetable production: a case of 512 farming households in Sichuan Provice." *Statistics & Information Forum* 27 (12):88-95.

Zhou, F. , and X. Xu. 2008. "A study on the pesticide use behaviour of non-hazard vegetable producers: the case of Nanjing." *On Economic Problems* 2008 (1):94-96.

Summary

Cooperation is vital for the sustainable development of human society. For the sustainable long-term development of human society, cooperation is essential in solving social dilemma situations ranging from local issues such as the management of local common pool resources to global issues such as climate changes. Understanding what affects cooperation and how to improve cooperation is crucial in this respect.

This thesis examines factors that affect cooperation among individuals in the context of China. It starts with an assessment of an eco-certification system for agricultural products, where cooperation among a larger number of farmers is crucial for its success. Upon finding that free-riding incentives are behind the ineffectiveness of the eco-certification system in reducing agrochemical use, this thesis continues to explore how various contextual and institutional factors affect cooperation. First, it looks at the role of contextual factors, specifically, the role of resource scarcity in shaping cooperation in the context of irrigation agriculture. Then, it moves on to the role of punishment and reward institutions in improving cooperation with particular attention to people's institutional preferences. And last, this thesis examines the effect of leading-by-example and leadership legitimacy on promoting cooperation. In Chapter 1, we provide an overview of the thesis, motivate each of the research questions and introduce the research methodologies used in the core chapters.

In Chapter 2, we examine the effectiveness of food certification in reducing agrochemical consumption in China. Specially, we use panel data from 4,830 different households in six provinces coving the period of 2005-2013 to test whether the adoption of certified food production reduces the use of chemical fertilizers and pesticide in China. On average, we do not find evidence support the hypothesis that adopting certified food production reduces farmers' agrochemical consumption. The effects are heterogeneous across villages, but the heterogeneous effects show no clear pattern that is consistent with the requirement of different types of certification. Our results are robust to the use of alternative panel structure and certification indicators. We find evidence suggesting that lack of knowledge about certification among farmers, weak inspection and the monetary incentives (price premium for certified products) for farmers to defect may explain why eco-certifications largely fails to reduce agro-chemical use. We interpret the role of monetary incentives in explaining the ineffectiveness of food certification in terms of a lack of cooperation among farmers.

With this case of failed cooperation among farmers in mind, in the rest of the thesis. we examine what affect cooperation and how to improve cooperation. In Chapter 3, we study the impact of long-term exposure to resource scarcity on cooperation among farmers using data from a household survey and a lab-in-the-field experiment with 312 male subjects in rural northwestern China. The unique historically formed irrigation water quota system allows us to measure exogenous variation in scarcity level of available water resource within an otherwise homogenous region. We find that water scarcity improves irrigation management in term of both the irrigation-related activities and their outcomes: people living in more water scarce villages are more likely to coordinate in crop choices, more likely to keep local canals clean and higher self-reported quality of canals. More important, we find that the impact of water scarcity goes beyond irrigation-related activities. People in villages with higher level of water scarcity also make significantly higher contributions in the public goods game. This result suggests that water scarcity strengthens norms of cooperation within rural communities.

In Chapter 4, we study how the people's institutional preferences interact with assigned institutions in the context of public goods games and what factors are behind people's institutional preferences using a lab-in-field experiment with 312 male subjects in rural northwestern China. We find that subjects have stronger preference for the reward institution over the punishment institution. But whether subjects' preferred institution matches the exogenously assigned institution or not does not have significant impacts on their contributions in the public goods game. Moreover, we find that subjects who prefer punishment tend to be free-riders. This finding makes the preference for the punishment puzzling and intriguing. Neither strategic concerns nor game history can fully explain why some people prefer punishment and the negative relationship between the preference for punishment and contributions in PGGs. We further find that the there is a robust relationship between the preference for the punishment institution and certain "efficiency-reducing" or "anti-social" social preferences profiles.

In Chapter 5, we study the role of leadership and leadership legitimacy in promoting cooperation using a lab experiment with 272 college students in China. We use a special experimental design to select leaders from the procedure and manipulate leadership legitimacy perception though manipulating the information provided to the leaders. We find that having a leader improves cooperation in the public goods game. The increase of group contributions is induced by the leader's contributions. Being a leader in makes leaders increase their contributions, and followers reciprocate but to a slightly smaller extent and thus they harvest

the gains from increased group contributions. The perception of leadership legitimacy does not have additional impacts on the leaders' contribution level and the group contribution level. But we find evidence suggesting that leadership legitimacy influences how leaders make their contributions decisions and update their beliefs in the repeated public goods game. It seems that only "legitimate" leaders show the strategic use of "leading-by-example", not "appointed" leaders, whose higher contributions as leaders are simply a reaction to the sequential move game structure where their behavior can be observed by others.

Finally, in Chapter 6, we conclude the thesis by discussing the policy implications, limitations and recommendations for future research.

Acknowledgement

This small book concludes the first step of a journey that started five years ago when I stumbled upon the job post of the JSTP project. Without knowing where exactly Wageningen was, I signed up for the opportunity. I am grateful that Nico and Tu Qin saw potentials in me in the hour-long interview and decided to offer me the position, which allowed me to pursue an academic career. I am deeply grateful to my supervisors, Nico Heerink, Qin Tu and Erwin Bulte, who not only guided my research, but also offered me freedom to do what I saw fit when my research interest deviated from the original project. In a way, I cherish the later even more as it allowed me to explore new fields and eventually helped me find the direction of my future career. In particular, I would like to thank Qin Tu. Besides acting as my supervisor, he is also a generous host to my frequent and often lengthy visits to Beijing, providing me with great office space, convenient accommodation and countless meals and tea. I am grateful to all my coauthors, Nico, Qin, Xiaojun, Jianying, Shuqin and Niccolo. I won't be here without you and I look forward to our continued collaboration.

My gratitude also goes to all the other people who contributed to the completion of this thesis. Hundreds of rural residents in Minle County and students in Beijing Normal University participated our survey and experiments. My research would be impossible without their participation and cooperation. I am grateful to Haoran He and Yuling Han, who helped me with the economics lab in Beijing Normal University. I am deeply grateful to the survey team that carried out the field work in Minle. They did an excellent job despite the cold weather. I am particularly in debt to Wei Qu, who helped oversee and coordinate our field work. Without her experience and local connections, our work in Minle would certainly not go as smoothly.

The life of a PhD student is not just about work. I would like to thank Jinghui, Yu, Min, Song, and many other Chinese friends. The casual conversations and warm dinners we had together brought the homely feeling to me in a land far far away from home. I am also grateful to Maria, Koen, Niccolo, Paul, Jan, Francesco, Aussi, Yan, Landry, Haki, Ian and all the other fellow students and colleagues and friends at DEC for the joyful and memorable time I had in the past few years.

Finally, I owe my family a debt of gratitude. I would not be able to finish my PhD without them. I am thankful for my parents who gave me all the freedom to pursue what I liked and had always been understanding and supportive of my choices. My warmest thanks belong to

my wife and best friend, Huimin. She accompanied me from Beijing to Wageningen, provided me with constant love and care, tolerated my moods, childishness and laziness, saw my best and worst and made my life absolutely wonderful. This thesis belongs to her as much as it belongs to me.

Zihan Nie
Wageningen School of Social Sciences (WASS)
Completed Training and Supervision Plan

Name of the learning activity	Department/Institute	Year	ECTS*
A) Project related competences			
Advanced Microeconomics, ECH 32306	WUR	2013	6
Behavioral and Experimental Economics, ECH51306	WUR	2014	6
Central Themes in Development Economics, DEC 30306	WUR	2014	6
PhD Research Proposal writing	WUR	2013	4
B) General research related competences			
WASS Introduction Course	WASS	2014	1
Research Methodology: From Topic to Proposal	WASS	2014	4
The Essentials of Scientific Writing & Presenting	WGS	2016	1.2
Scientific Writing	WGS	2017	1.8
C) Career related competences/personal development			
'Does certified food production reduce Agro-chemicals use in China'	7th CAER-IFPRI Annual Conference, Zhejiang A&F University, IFPRI, China Agricultural University	2015	1
'Authentic leadership, leaders' legitimacy, and group provision to the public good: an Experiment on what motivates leaders to act as role models'	IAREP/SABE Conference, Wageningen University	2016	1
'Water scarity and cooperation: evidence from rural China'	12th Nordic Conference on Behavioural and Experimental Economics, University of Gothenburg	2017	1
	9th CAER-IFPRI Annual International Conference, China Agricultural University	2017	
'Legitimacy, leading by example and public goods provision: evidence from a sequential public goods game'	Beijing Normal University Conference of Experimental Economics	2017	1
'Resource scarity and cooperation: evidence fron an irrigation system in western China'	Seminar, University of Gothenburg	2018	1
Total			**35**

*One credit according to ECTS is on average equivalent to 28 hours of study load

The research described in this thesis was financially supported by the Dutch Organization for Scientific Research (NWO), project no. 833.13.003.

Printed by: Digiforce || ProefschriftMaken