途上国における貧困削減と制度・市場・政策
比較経済発展論の試み
Poverty Reduction, Institutions, Markets, and Policies in Developing Countries:
Toward a Theory of Comparative Economic Development

PRIMCED Discussion Paper Series, No. 66

Incentives and Social Preferences: Experimental Evidence from a Seemingly Inefficient Traditional Labor Contract?

Jun Goto, Yasuyuki Sawada, Takeshi Aida,

and Keitaro Aoyagi

March 2015

Hitotsubashi University

Research Project **PRIMCED**
Institute of Economic Research
Hitotsubashi University
2-1 Naka, Kunitatchi Tokyo, 186-8601 Japan
http://www.ier.hit-u.ac.jp/primced/e-index.html

Incentives and Social Preferences:
Experimental Evidence from a Seemingly Inefficient Traditional Labor Contract*

Jun Goto
Hitotsubashi University

Yasuyuki Sawada[1]
University of Tokyo

Takeshi Aida
Graduate National Research Institute for Policy Studies

Keitaro Aoyagi
University of Tokyo

March 2015

Abstract

This paper investigates the interplay between economic incentives and social norms in formulating rice planting contracts in the Philippines. In our study area, despite the potential for pervasive opportunistic behaviors by workers, a fixed-wage (FW) contract has been dominant for rice planting. To account for the use of this seemingly inefficient contractual arrangement, we adopt a hybrid experimental method of framed field experiments by randomized controlled trials (RCT), in which we randomly assign three distinct labor contracts—FW, individual piece rate (IPR), and group piece rate (GPR)—and artefactual field experiments to elicit social preference parameters. Through analyses of individual workers' performance data from framed field experiments and data on social preferences elicited by artefactual field experiments, three main empirical findings emerge. First, our basic results show the positive incentive effects in IPR and, equivalently, moral hazard problems in FW, which are consistent with standard theoretical implications. Second, non-monetary incentives seem to play a significant role under FW: while social preferences such as altruism and guilt aversion play an important role in stimulating incentives under FW, introducing monetary incentives crowds out such intrinsic motivations, and other non-monetary factors such as positive peer effects significantly enhance incentives under a FW contract. Finally, as alternative hypotheses, our empirical results are not necessarily consistent with the hypothesis of the interlinked contract of labor and credit transactions in mitigating moral hazard problems, the optimality of FW contract under large effort measurement errors, and the intertemporal incentives arising from performance-based contract renewal probabilities. Hence, considering the interplay of intrinsic motivations and monetary incentives as well as the monetary costs of mitigating moral hazard and free-riding problems through IPR, we may conclude that seemingly perverse traditional contractual arrangements might be socially efficient.

Keywords: Randomized controlled trials, incentives, social preferences, peer effect, labor contract, field experiments

JEL: D03, C93, D22, C91

---

* The authors would like to thank Jonna P. Estudillo, Thelma B. Estera, Hitoshi Matsushima, Tetsuji Okazaki, Keijiro Otsuka, Hideo Owan, and Tomomi Tanaka for their encouragement concerning this project and Oriana Bandiera, Alan de Brauw, Alain de Janvry, Geary Fields, Ronald Findley, Tomoki Fujii, Michael Carter, Julen Esteban-Pretel, Marcel Fafchamps, Drew Griffen, Hal Hill, Hide Ichimura, Shingo Ishiguro, Michihiro Kandori, Hisaki Kono, Takashi Kurosaki, Jungmin Lee, Momoe Makino, Chiaki Moriguchi, Akihiko Matsui, Alistair Munro, Valerie Mueller, Takeshi Murooka, Masayuki Nakabayashi, Ryo Nakajima, Peter Orazem, Keijiro Otsuka, Albert Park, Mark Rosenzweig, Betty Sadoulet, Tatsuyoshi Saijo, Paul T. Schultz, Mandy Shen, Andrew Sim, Tetsushi Sonobe, Katsuya Takii, Eric Thorbecke, Angelino Viceisza, Eric Weese, David Yanagizawa-Drott, Yasutora Watanabe, Jane Zhang, Junsen Zhang, and the participants at the presentations at MIEDC 2012, NEUDC 2011, IFPRI, IRRI, 2012 and 2013 Asia-Pacific ESA, Tokyo, ICAE 2012, and at the Universities of ANU, GRIPS, Kyoto, SMU, UC Davis, London, Tokyo, Waseda and Yokohama National for their useful comments on earlier drafts. The usual disclaimers apply.

[1] Corresponding Author: Faculty of Economics, University of Tokyo; Email: sawada@e.u-tokyo.ac.jp; Tel: +81-3-5841-5572; Address: 7-3-1 Bunkyo-ku, Tokyo 113-8657 Japan

1. Introduction

The fixed-wage (FW) contract for rice planters has been the dominant contract in Central Luzon province in the Philippines. Since rice planting supervisors, called *kabisilyas*, can observe individual work–effort outcomes accurately to some extent (Picture 1), such a FW contract with low intensity incentives is likely to be sub-optimal, generating a moral hazard problem. In this paper, we aim to solve this puzzle of the real-world perverse contract by a hybrid method of framed field experiments by randomized controlled trials (RCT), in which we randomly assign different labor contracts, and artefactual field experiments to elicit social preference parameters.[2] By doing so, we believe that we will be able to better understand the economic rationale of the seemingly inefficient existing agricultural labor contracts which, are crucial in setting rural development and poverty alleviation strategies (Hayami and Otsuka, 1993; Roumasset and Lee, 2007).

Generally speaking, there are two theoretical ways to argue the optimality of FW contracts. First, in line with the arguments by Bowles (2008), intrinsic motivations may prevent workers from displaying opportunistic behavior even under a FW contract. The importance of the interaction between intrinsic motivations and monetary incentives in the workplace has been incorporated in economic theory (Kreps, 1997; Benabou and Tirole, 2003; Itoh, 2004; List and Rasul, 2011; Bowles and Polania-Reyes, 2012, Koszegi, 2014). This attempt extends standard agency models to take into account other-regarding preferences in mitigating moral hazard problems (Kandel and Lazear, 1992; Rotemberg, 1994; List and Rasul, 2011).

Second, with standard theories, we may be able to explain the optimality of FW contracts. Since the optimal intensity of the incentive condition in a standard agency model tells us that the slope of the performance-based wage compensation scheme is a negative function of the observability or

[2] While on each day of the field experiments, we worked with basically the same set of workers, randomly assigning different contracts in the morning and afternoon sessions, the participating rice planters differed across days. This setting provides within-subject variations as well as between-subject variations in our data.

precision of worker performance (Milgrom and Roberts, 1992), with sufficiently large measurement errors of worker performance, it is optimal for the principal to set a FW contract. Secondly, the optimality of FW contracts can be shown in a multitask principal–agent model in which the performance of one of the indispensable tasks is unmeasurable or unverifiable (Proposition 1 of Holmstrom and Milgrom, 1991). While it may be said theoretically that rice planting is a multitask work of planting speed and quality, in reality, rice planting tasks are rather simple and are difficult to manipulate (Umekage at al., 1965).[3] Third, the intertemporal incentive for contract renewal can be consistent with the optimality of a FW contract. In a situation in which a supervisor has a right to decide the contract renewal of workers, contract renewal probability can be a positive function of work performance, which can be observed to some extent. Then, while the instantaneous wage is fixed, the expected wage becomes effectively the piece rate, mitigating moral hazard. This possibility is in line of theoretical models of effort determination under FW through termination contracts (Shapiro and Stiglitz, 1984; List and Rasul, 2011,). Finally, the FW contract can be optimal if it is interlinked with other contracts, such as a credit contract. This is simply because by changing the terms and amounts of loans to workers, lenders can induce more efforts, solving moral hazard problems in labor contracts (Braverman and Stiglitz, 1982).

As to these two ways of explaining the optimality of FW contracts, existing empirical studies using personnel data indicate that “monetary incentives” as well as “social incentives,” which are an important part of other-regarding preferences, can increase workers’ productivity and mitigate agency problems (Lazear, 2000; Shearer, 2004; Mas and Moretti, 2009; Bandiera et al., 2010). However, what

---

[3] Umekage et al. (1966) carried out experiments to identify the effect of various transplanting methods on the growth habits of rice plants by comparing two methods: planting rice seedlings in a horizontal direction and in a vertical direction (which is the normal method). They found that the difference in the yields between the plots was not significant at the 5% level. Moreover, according to the *kabisilyas*, not only the speed but also the quality of planted rice can be easily observed and verified by checking the misalignment and lodging of seedlings. Hence, we believe that multitask problems are not necessarily serious in our experiments. In addition, in our experiment instructions, we emphasized that we will not make experiment payments for lodged seedlings to mitigate the multitask problems. The *kabisilyas* told us that no systematic negative harvest effects arise from the possible lodging of seedlings.

sort of social motivation shapes people's behavior and how this motivation interacts with monetary incentives are still under-investigated in the literature. The objective of this paper is to bridge this gap in the existing empirical studies in two ways. First, we investigate the interplay between economic incentives and intrinsic motivations. More precisely, we examine whether monetary incentives crowd out intrinsic motivation by a hybrid method of artefacutual and framed field experiments (Harrison and List, 2004; Levitt and List, 2009). Second, we test alternative hypotheses to the crowding out hypothesis, including the effort measurement error, intertemporal incentives, and interlinked contract hypotheses. To these purposes, we combine individual performance data from rice planting framed field experiments in the Central Luzon province of the Philippines, where we randomly assign three distinct labor contracts: the FW contract as well as the individual piece rate (IPR) and group piece rate (GPR) contracts, with data on social preferences constructed from the results of the artefactual field experiments. These hybrid experimental data enable us to specify the underlying mechanisms for mitigating agency problems. Since the number of empirical papers employing field experiments in labor economics is still thin (List and Rasul, 2011), we believe we make an important contribution to the literature.

Three main findings emerge from our analysis. First, our basic results show positive incentive effects in IPR and moral hazard problems in FW, which are consistent with the standard theoretical implications as well as the empirical findings of Foster and Rosenzweig (1994), Lazear (2000), and Shearer (2004). Second, non-monetary incentives seem to play a significant role under FW. While social preferences such as altruism and guilt aversion play an important role in stimulating incentives under FW, introducing monetary incentives crowds out such intrinsic motivations. This is a rare real-world finding that confirms evidence from the laboratory experiments reviewed by Bowles (2008). In addition, other non-monetary factors such as positive peer effects significantly enhance incentives under the FW contract, which is consistent with the findings of Mas and Moretti (2009) and Bandiera et al. (2010). Overall, significant monetary costs to introduce monetary incentives can explain the optimality of the FW contract. Finally, as alternative hypotheses, our empirical results are not necessarily supportive of the hypothesis of the interlinked contract of labor and credit transactions in mitigating moral hazard problems, the optimality of the FW contract under large measurement errors, or intertemporal incentives arising from performance-based contract renewal probabilities. In sum, while social preferences that appear as altruism or inequality aversion seem to mitigate agency problems in the workplace, these effects might have been crowded out by monetary incentives. Hence, considering the interplay of intrinsic motivations and monetary incentives as well as the monetary costs of mitigating moral hazard and free-riding problems through IPR, we may conclude that these seemingly perverse traditional contractual arrangements might be socially efficient.

The rest of this paper is organized as follows. Section 2 describes the context of our study and our framed and artefactual field experiments. In Section 3, we show our data with summary statistics. The empirical results are presented in Section 4. Section 5 provides concluding remarks.

2. Experimental Design

2.1 Rice Planting in the Study Villages

In the Central Luzon province in the Philippines, labor to transplant seedlings manually is traditionally supplied by agricultural laborers and members of tenant families from local villages at the beginning of each regular crop season. Since farmers do not want to miss the appropriate timing of rice planting when irrigation water becomes available, during peak planting season, demand for hired rice planting labor is concentrated around the same time in each village. To allocate planting labor effectively, loosely tied labor groups for rice planting are formed by supervisors called *kabisilyas*, literally meaning "boss" in Tagalog. Figure 1 illustrates the structure of the rice planting organization and payment scheme. First, a farmer outsources rice planting to a *kabisilya* at a cost of 2,200 PHP/ha. The total amount of payment from a farmer to *kabisilya* depends only on the total contracted size of the paddy field. After the *kabisilya* subtracts her fixed margin from the total revenue,[4] that is, 2,200 PHP multiplied by the total planting area, the rest of the amount is equally divided among the planting workers. While, theoretically speaking, the number of planting workers is endogenously determined, the prevailing norm in the area has been to recruit an appropriate number of workers so that each daily payment becomes 100–120 PHP per worker. In this way, workers receive a fixed wage regardless of their performance.[5] During the peak planting season, workers are asked to work in the paddy fields almost every day, and they typically plant one to two hectares per day together with their fellow workers. Workers are not permitted to plant in other landowners' plots before finishing their duties in a given day. In a paddy field, workers voluntarily form subgroups consisting of five to seven members and take responsibility for planting in their assigned plots. The composition of these subgroups usually changes daily. Accordingly, the choice variables for planting workers are their selection of coworkers and the level of effort they put into planting.

2.2 The Hybrid Experiment

In this study, we adopt a hybrid method of two types of field experiments: framed field experiments and artefactual field experiments (Harrison and List, 2004; Levitt and List, 2009). The former type changes the rice planting labor contract, while in the latter type, laborers participate in a series of carefully designed laboratory economic experiments.

2.2.1 Framed Field Experiments

In our framed field experiments, all workers are asked to take part in one of three randomly assigned labor contracts: FW, IPR, and GPR. While group formation for actual rice planting has been

---

[4] The margin is 100–150 PHP.

[5] Strictly speaking, the reward level for each worker relies on the number of planters. However, the ratio of the number of recruited workers in a day to the total contracted area hardly varies. Hence, the prevailing type of contract is effectively a pure fixed-wage contract.

based on self-selection mechanisms, we also introduce random group assignments. Hence, we have a total of six arms in our randomized controlled trials: all the combinations of the three assigned contracts and two methods of group formation.[6] In each day, we work with the same set of workers, randomly assigning different contracts in the morning and afternoon sessions. This setting provides us with within-subject variations in our data. However, the participating rice planters differ each day. Hence, strictly speaking, we utilize a combination of within- and between-subject variations.

The Three Wage Contracts

Let l_{ij} be the output level chosen by worker *i* in a group *j*, which is measured by the average length of line planted every 10 minutes for 30 minutes.[7] Each day, at the beginning of the morning and afternoon planting sessions, we select a contractual arrangement randomly from the three types.[8] Suppose that the observed outcome is decomposed into the true effort level of worker *i*, e_{ij}, and the well-behaved measurement error, ε_{ij}: $l_{ij} = e_{ij} + \varepsilon_{ij}$, where only l_{ij} can be observable and used for a performance-based contract. The prevailing rice planting contract is a FW contract under which the amount of the fixed wage, w_{ij}^{f}, for a worker *i* belonging to subgroup *j* is simply $w_{ij}^{f} = F$. Under the IPR contract, a wage rate, w_{ij}^{p}, is composed of two parts, a fixed component and an incentive component that is proportional to individual productivity: $w_{ij}^{p} = F + \alpha\, l_{ij}$, where we impose that $0 \leq \alpha\, l_{ij} \leq F$. Hence, the reward amount is set at the level at which the most productive worker can be paid twice the fixed rate. Under the GPR contract, a wage rate, w_{ij}^{g}, is also composed of two parts, a fixed compensation part and an incentive component that is proportional to the average productivity of the group rather than each individual: $w_{ij}^{g} = F + \alpha \sum_{i\in j} l_{ij}/n_j$, where we impose that $0 \leq \alpha \sum_{i\in j} l_{ij}/n_j \leq F$ and n_j is the number of workers in group *j*. We set the fixed wage, *F*, equal to 50 PHP and the intensity of the

---

[6] During the first half of the planting season, each planting group is formed randomly by us; during the second half, workers form their own planting groups.

[7] While we measure the length for 30 minutes, we construct three ten-minute observations for each worker.

[8] To avoid confounding the effects of each treatment with other factors, the compensation schemes for the *kabisilyas* remain unchanged.

incentive parameter, α, equal to two PHP.

It is straightforward to show that the marginal monetary benefits of putting additional effort into FW, IPR, and GPR contracts are $\frac{\partial w_{ij}^{f}}{\partial e_{ij}} = 0$, $\frac{\partial w_{ij}^{p}}{\partial e_{ij}} = \alpha$, and $\frac{\partial w_{ij}^{g}}{\partial e_{ij}} = \frac{\alpha}{n_j}$, respectively. Hence, while the marginal monetary benefit is highest in IPR, followed by GPR, the marginal monetary benefit is zero under FW. Hence, in terms of monetary incentives (Lazear, 2000), the observed gap in worker performance between IPR and FW reflects the moral hazard problem, and the difference in the individual worker outcomes between IPR and GPR exhibits the free-riding problem.

Randomized Team Formation

The team formation procedure in our experiment is structured in two ways: randomly formed groups and self-selected groups. In both groupings, teams are newly formed for each rice planting session. The basic team is a group of five workers, but some adjustments are made based on the total number of workers who show up to the field.[9] In one setting, all planting workers recruited by a *kabisilya* at a given paddy field are randomly assigned to teams.[10] A plot is randomly assigned to each team by the experimenters. We refer to this arrangement as "randomly assigned groups." In the other setting, workers are asked to organize their planting teams by themselves in each session. We call this arrangement the "self-selected groups."[11] Two potential factors drive team composition in the latter setting. First, workers may prefer to match with high-ability colleagues to complete planting in the assigned rice fields as quickly as possible. In this case, all else being equal, we would expect assortative matching patterns. Alternatively, workers may prefer to form teams with socially connected members because they derive direct utility or can achieve cooperation in a self-enforcing manner (Bandiera et al.,

---

[9] The upper and lower bounds of the number of teams and their members are set because of the limited availability of plots for experiments and the need to secure accuracy in measuring productivity by enumerators.

[10] The number of workers recruited by *kabisilyas* depends on the location of the paddy field and the total contracted area.

[11] Theoretically, self-selection by workers represents a coalition-proof Nash equilibrium (Bandiera et al., 2013).

2005, 2010).

Experiment Procedures

In each experiment, a contract type is randomly chosen and assigned by experimenters at the beginning of each experiment. After an announcement of the selected contract, randomly assigned groups or self-selected groups are formed. This experiment is carried out twice in each day, in the morning and the afternoon. In order to avoid anticipation effects, the workers and *kabisilya* are notified of the selected contract right before each rice planting session. Enumerators are also randomly assigned to each subgroup to measure workers' individual productivity and are required to record productivities by tracing each worker's planting line by hand on carefully designed reporting sheets for the first 30 minutes of each planting session.[12] Rewards for individual workers in each experiment are paid immediately at the paddy field before starting the next round of the experiment. The actual payments for each worker are kept strictly confidential so that other factors, such as self-image construction, are carefully controlled.

2.3 Artefactual field Experiments

To examine the optimality of FW contracts, we follow the arguments by Bowles (2008) to hypothesize that intrinsic motivations may prevent workers from employing opportunistic behavior even under a FW contract. Such intrinsic motivations may arise from Filipino community norms such as *pakikisama*, which literally means "going along well with others" (Hayami and Kikuchi, 2000). To elicit each worker's social and individual preference parameters, we conducted four standard artefactual field experiments with the same rice planting workers on one day during the rice planting

[12] During the pilot experiments, this method was validated by comparing it with alternative methods. First, we used video camera to recode each workers' activities in a given plot; second, we asked workers to put high- and low-quality passometers on their waists during rice planting. Since videotaping and decoding are too costly and we cannot achieve consistent measurements using different passometers, we decided not to use them.

season: the dictator game, the ultimatum game, the public goods game with monitoring and disapproval messages, and the risk game. At the end of the experiments, each subject received a real monetary payoff based on a randomly selected outcome of the experiments together with a fixed show-up fee of 100 PHP.[13]

The Dictator Game

We conducted a standard dictator game in an anonymous setting to elicit altruism toward a co-worker under the same *kabisilya* (Camerer and Fehr, 2004; Levittt and List, 2007). Each player was given an envelope with 10 coins worth 10 PHP each.[14] They then decided how many coins to put into their partner's envelope. However, the players were not informed about identity of their partners.

The Ultimatum Game to Elicit Guilt Aversion and Nonenviousness

The first move in the ultimatum game is the same as the dictator game, but a second move is added in this game (Roth, Prasnikar, Okuno-Fujiwara, and Zamir, 1991): each sender or "proposer" is informed beforehand about the rule that their keeping and sending amounts materialize only if the "responder" accepts their offer; if the responder rejects the offered amount in the first move, the proposer and responder both receive nothing. After each player decided how many coins to allocate to his or her partner, they were also asked which offer they would accept as a responder using the strategy method. The observed outcomes in the ultimatum game can be explained using standard equilibrium concepts applied to the inequality aversion utility of Fehr and Schmidt (1999), in which players care about their own payoffs and the difference between their payoffs and those of others. Specifically, we quantify two aspects of inequality aversion. First, since the proposers' offer amounts in the ultimatum

[13] Detailed descriptions of artefactual field experiments and actual implementation procedures are available from the corresponding author upon request.
[14] 100 PHP is equivalent to the daily wage rate for manual labor. Each subject also received a show-up fee of 150 PHP and a free lunch. Payments for artefactual field experiments were made for a randomly chosen experiment from all those conducted.

game depend on the guilt weight of Camerer (2003, pp. 102–103) in the inequality aversion preference, we interpret the sending amounts in the ultimatum game as an observed level of "guilt aversion." Since altruism can also affect the sending amount, we analyze the data after subtracting the sending amount in the dictator game from that in ultimatum game. Second, the minimum acceptance level of a responder, which is based on the inequality aversion model of Fehr and Schmidt (1999), is a "negative" function of inequality aversion or the envy weight parameter of Camerer (2003). Hence, we take the gap between the maximum value, 10 coins worth 10 PHP, and the minimum acceptance level as a responder to construct a measure of "nonenviousness."

The Public Goods Game with Monitoring and Disapproval Messages

We conducted a public goods game with monitoring and disapproval messages based on Carpenter and Williams (2014). In this setting, workers formed anonymous groups of four that stayed together for the entire experiment. At the beginning of the first round, each player was given an envelope with 10 coins of 10 PHP. They were then asked how many coins to contribute to the group project, keeping the rest for themselves in a non-cooperative setting. The total group contribution was doubled and redistributed equally to all four members.

After this contribution stage, we showed each worker her gross income and total contribution. The workers were then given a chance to monitor the contribution amounts of each group member anonymously by paying one PHP. If a worker monitored other players' contributions, he/she could send messages of disapproval (unhappy faces) to other individuals in the group for an additional one PHP per message. We continued these contribution and monitoring stages for three rounds. However, the workers were not allowed to send any messages in the third round, in which there was no room to improve other players' contribution levels.

Since a pure-strategy Nash equilibrium of the public goods game is a zero contribution by all subjects, existing studies such as Camerer and Fehr (2004) and Levitt and List (2007) have found that observed contributions reflect a type of social preference because of *reciprocal expected cooperation*.

Each worker's *propensity for monitoring* and *propensity for sanction* are captured by the raw monitoring frequency and the number of disapproval messages sent, respectively (Carpenter and Seki, 2005; Carpenter and Williams, 2014).

The Risk Game

In addition to the experiments used to elicit social preference parameters, we also conducted the risky investment game of Schechter (2007) to measure each worker's risk preference. Each subject received an envelope with 10 coins of 10 PHP and decided how many coins to invest in a risky investment game and how many coins to take home without investments. Then he/she flipped a coin: if it was heads, their invested amount was doubled, but if it was tails, the subject lost the entire invested amount. The invested amount in this game thus represents each worker's individual risk tolerance level.

3. The Data

Our two study villages, G village and M village, are located in Nueva Ecija Province in the Central Luzon region of the Philippines. These villages were studied intensively through projects supported by the Social Science Division of the International Rice Research Institute (Otsuka, 1991; Otsuka, Chuma, and Hayami, 1993; David, Cordova, and Otsuka, 1994; Estudillo, Sawada, and Hossain, 2005; Estudillo, Sawada, and Otsuka, 2008, 2009). We conduct framed and artefactual field experiments for 10 and two days, respectively, during the dry planting season in 2011. Our sample is composed of 120 workers, 58 of whom were from G village and 62 from M village. Without prior notice to workers and enumerators, at each morning and afternoon session, we randomly introduced one of the three distinct monetary incentives—the FW, IPR, and GPR contracts—under randomly assigned or self-selected groups. We obtained a total of 1,884 observations from 28 field experiment sessions: 15 sessions for G village and 13 sessions for M village over 10 days (Figure 2).

We conducted three additional surveys: a short questionnaire survey for workers after each field experiment session to capture session-specific information such as subjective health conditions; second, a household survey to collect information on the individual and household characteristics of each worker; and third, a survey of *kabisilyas* after the final field experiment to evaluate the subjective performance of each worker.

Toward the end of the field experiments, we spent two days, one in each village, conducting artefactual field experiments: we invited the rice planting workers who participated in the rice planting experiments to participate in a series of laboratory experiments (Figure 2). Each day, the experiment took four to five hours. The total number of participants was 108: 53 from G village and 55 from M village.

3.1 Data from Framed Field Experiments

Table 1 shows the descriptive statistics of individual rice planting performance in the field experiments, the conditions of the field experiments, and the individual characteristics collected via the surveys. Panel A shows the pro-incentive effects in the IPR contract: on average, workers could achieve the highest productivity under IPR (29.97 m/10 min), followed by GPR (29.75 m/10 min) and FW (26.08 m/10 min). While the average productivity difference between FW and GPR is statistically significant at the 1% level, the mean difference between GPR and IPR is not significant. Figure 3 compares the cumulative distribution functions (CDFs) of individual rice planting productivity in the three contracts. We can easily verify that the CDF of the IPR contract dominates that of the FW contract. The two-sample Kolmogorov–Smirnov tests of the equality of distributions reject equality between these two CDFs at the 1% level of statistical significance. In contrast, the equality hypothesis of the CDFs between the IPR and GPR contracts can be rejected only at the 10% significance level. This is not necessarily surprising because people often react to the mere existence of incentives rather than to their extent (Bowles and Polania-Reyes, 2012).

To illustrate the overall pattern on peer effects within each rice planting team, Figure 4 shows the correlation between individual productivity and other members' average productivity on the same team. We find that individual productivity is positively correlated with the others' productivity, suggesting that there is a strong positive peer effect.

Panel B of Table 1 presents the subjective relationships among workers within each rice planting team. The table shows that the average years of acquaintance with the other group members is 10.64 years. Moreover, 39.38% of the group members attend the same church, where regular services are held on Sundays.

As shown in panel C of Table 1, which summarizes the individual characteristics, the average age of the workers is 35.6 years. Subjects tend to be landless agricultural workers; nearly all received advanced payments or, equivalently, credit from their *kabisilya* in every planting season. The average amount of outstanding advanced payment was 625 PHP. As to the pattern of self-reported relationships with their *kabisilya*, more than 38% of surveyed workers were family members or relatives of their *kabisilya*. This is not unusual in Central Luzon, where most workers come from the same *barangay* (village) as their *kabisilya*, who usually recruits workers through her social networks in the village.[15]

3.2 Data from the Artefactual Field Experiments

Panel A of Table 2 shows the summary statistics of the results of the dictator game and the ultimatum game. The average sending amounts in both games are consistent with the results reported in previous studies such as Camerer and Fehr (2004) and Levitt and List (2007). The average sending amount in the ultimatum game is larger than that in the dictator game, which is also consistent with the previous studies (Forsythe et al., 1994). The difference in the sending amount between the ultimatum and dictator games can be interpreted as the degree of "pure" guilt aversion. Figures 5 and 6 show the histograms of the sending amounts of the dictator and ultimatum games, respectively. While in the

[15] The anthropological literature describes a village or *barangay* in Tagalog as a traditional community group composed mainly of kin or extended family (Kerkvliet, 1990).

dictator game, most workers send less than 50 PHP, there is a clear peak at 20 and 50 PHP in the ultimatum game. Figure 7 shows the histogram of the responders' minimum acceptance level in the ultimatum game. While the highest peak is at 0, the second-highest peak is at 50, implying the existence of the 50-50 norm (Andreoni and Bernheim, 2009).

Panel B of Table 2 shows the summary statistics of the results for the public goods game. The mean contribution over the rounds is 33.3 of the 100 PHP, indicating that the workers contributed one-third of their initial endowment. Figure 8 is a histogram of each worker's mean contribution level to public goods over all rounds: most workers contribute less than 40% of their initial endowment. As Table 2 shows, there is no clear trend in the contribution level over time, which is similar to the results of Carpenter and Williams (2014) but different from the standard voluntary contribution games without monitoring and disapproval messages (Ledyard, 1995). As shown in Table 2, the monitoring intensity decreases slightly as the round proceeds, which is consistent with Carpenter and Williams (2014). The number of disapproval messages sent by each worker increases from the first round to the second round.[16]

Panel C of Table 2 shows the result of the risk game. The mean invested amount is 37.7 PHP and all workers invest at least 10 PHP. Since the expected net return from this experiment is zero, the investment amount represents each worker's attitude toward risks. Figure 9 shows the histogram of the risk game, in which we can see that most investment amounts are less than or equal to 50 PHP.

4. Empirical Analysis

4.1 Econometric Framework

We follow Lazear (2000), Shearer (2004), Bandiera et al. (2005, 2010), and Mas and Moretti

[16] Note that workers were not allowed to send any messages in the final round.

(2009) to postulate the following econometric model of rice planters' observed work performance:

$$l_{ijt} = \alpha_t + \lambda^{FW} FW_t + \lambda^{GPR} GPR_t + S_{ijt}\delta + \mu\bar{P}_{-ijt} + \pi R_t + X_{ijt}\gamma + \varepsilon_{ijt}, \qquad (1)$$

where l_{ijt} is the productivity of worker *i* on team *j* measured in terms of the average planting length in meters per 10 minutes in the *t*-th round; α_t is the round (time) effect; FW_t and GPR_t are the dummy variables for the FW and GPR contracts, respectively, where the IPR contract is taken as the default category; $\bar{P}_{-ijt}$ is the average ability of the other workers on worker *i*'s team *j* as captured by the method of Mas and Moretti (2009); R_t is a group formation dummy variable that takes a value of one for a self-selected team and zero for a randomly assigned team; S_{ijt} is a set of social preferences measured by the artefactual field experiments; and X_{ijt} is a set of household characteristics.

Our parameter of interest is λ^{FW}, which captures the disincentive effect of shifting from IPR to FW on individual productivity. If $\lambda^{FW} < 0$, we may conclude that there exists moral hazard induced by the lack of monetary incentives. In addition, the finding that $\lambda^{GPR} < 0$ is consistent with the free-riding problem.

The estimated coefficients on $(S_{ijt}, \bar{P}_{-ijt}, R_t)$, ($\delta$', μ, π), capture the effects of social preference, peer effects, and self-selection, respectively, on worker incentives. Note that $\bar{P}_{-ijt}$ is defined as the average of other workers' permanent ability rather than their contemporaneous productivity in order to avoid the reflection problem of Manski (1993, 2000). To avoid the reflection problem, we adopt the two-step method developed by Mas and Moretti (2009).[17] If $\mu \neq 0$, a worker's individual behavior is affected by the mean of the group performance. This is the "endogenous effect" of Manski (1993). For the contract and team assignments, we select different wage contracts randomly and assign self-selected teams or randomly assigned teams exogenously. Hence, by our field

[17] In contrast, if we follow Hamilton, Nickerson, and Owan (2003) and use simple average productivity, the effect of average ability on individual productivity becomes uniformly larger. This indicates the problem of the upward endogeneity bias that arises from the reflection problem when we estimate peer effects.

experiments, we believe the OLS estimates of equation (1) can reveal the causal effects of contracts, peer effects, self-selection, and social preference on individual productivity fairly accurately.

4.2 The Ability Measure

The test of peer effects requires a measure of the permanent ability of worker i and her peers. To this purpose, we follow Mas and Moretti (2009) and Bandiera et al. (2010) to estimate a worker productivity model with worker fixed effects considering the fact that an individual's productivity may be affected by her co-worker composition. Specifically, we employ the following estimation model:

$$l_{ijt} = \beta_i + TEAM_{jt}\gamma + \epsilon_{ijt}^{a}, \quad (2)$$

where β_i is the individual fixed effects and $TEAM_{jt}$ is the dummy variables for the teams formed in each round. We construct each worker's permanent ability by estimating the worker fixed effects, β_i.

Using these permanent abilities, we can construct other group members' average ability $\bar{P}_{-ijt}$ and estimate Equation (1). However, we need to correct the standard errors in this two-step procedure. Following Mas and Moretti (2009), we employ the Bayesian parametric bootstrap method: first, we estimate Equation (2) to elicit each worker's permanent ability. Using the estimated coefficients, we construct 1,000 simulated datasets by randomly drawing β_i from the distribution, $N(\hat{\beta}, \hat{\Sigma})$, where $\hat{\beta}$ is the vector of the point estimates of β_i and $\hat{\Sigma}$ is the estimated variance–covariance matrix. Then, we estimate Equation (1) with these simulated datasets as well as the original dataset to obtain the coefficient on $\bar{P}_{-ijt}$, μ, for each dataset. The standard error of μ in Equation (1) is computed as $\sqrt{s_{\mu}^2 + \sigma_{\mu}^2}$, where s_{μ}^2 is the sampling variance of the original datasets and σ_{μ}^2 is the between-simulation variance, which is the variance obtained from simulated datasets.

4.3 Estimation Results

Table 3 shows the regression results for the estimation of Equation (1) with the full sample. In column (1), we estimate a simplified version of Equation (1) without social preference variables, S_{ijt}. The result shows that IPR in the default category generated higher productivity than FW, in which the average gap is around 3.6 meters per 10 minutes. However, the difference between IPR and GPR is not statistically significant. The former finding is consistent with the moral hazard problem, replicating the non-experimental results of Lazear (2000) and Foster and Rosenzweig (1994) and the experimental results of Shearer (2004), who determined the incentive effect generated by monetary incentives.

In column (2), we added a set of social preference variables elicited by artefactual field experiments and found that non-monetary incentives had a significant role. The difference in the offer amounts between the ultimatum and dictator games has positive and significant coefficients: people with the guilt aversion preference are likely to exert more effort in their rice planting work. In addition, nonenviousness, which is defined as the gap between the maximum value of the ultimatum game (i.e., 10 coins) and the selected values of the minimum acceptance level of responders in the game, has positive and significant coefficients. This suggests that nonenviousness stimulates positive work incentives in rice planting. Hence, our results indicate that individual social preferences affect own productivity, underlying intrinsic motivation.

In columns (3) and (4), we add variables related to peer effects, that is, the average ability of the other team members. The average ability of other the team members is positive and significant, indicating positive spillover effects among workers in the workplace. However, the self-selected group formation variable has a negative but insignificant effect. In addition, advanced payment has an insignificant effect on effort, a finding that is not necessarily supported by the theoretical results by Braverman and Stiglitz (1982) that interlinked contract of labor and credit transactions can mitigate moral hazard problems.

Table 4 shows the estimation results for productivity from Equation (1) separately for each incentive scheme. Social preferences stimulate significantly positive effects only under FW according to both the individual coefficient and joint significance test results. For the individual coefficients, the sending amount in the dictator game and the net offered amount in ultimatum game, which can be interpreted as degrees of altruism and guilt aversion, respectively, are positively correlated with individual productivity under the FW contract. In addition, the nonenviousness variable takes positive coefficients only for FW, indicating that nonenviousness mitigates disincentive effects or moral hazard in the FW contract. While in the case of FW, the joint F-test of the estimated coefficient of the social preference variables rejects the zero coefficients of these variables at 1%, the null hypothesis cannot be rejected in the IPR and GPR contracts. In addition, the non-monetary peer effects captured directly by the average ability of other team members have positive effects on individual productivity only in the FW contract. These results fits with the hypothesis that the extrinsic incentives of the individual piece rate would crowd out the intrinsic motivation of workers (Bowles and Polania-Reyes, 2012).

Blinder–Oaxaca Decomposition

To grasp the magnitude of the productivity effects of each component, we performed the Blinder–Oaxaca decomposition (Blinder, 1973; Oaxaca, 1973) for the average productivity difference between IPR and FW, which is 4.71 meters per 10 minutes (Table 5). Partly because of the construction of a *quasi* within-subject design, the endowment difference is not statistically different from zero.[18] Using the Blinder–Oaxaca decomposition, we can determine how much of this difference can be explained in terms of social preferences. The decomposition results, which are based on the estimation results shown in columns (4) and (5) of Table 4, are summarized in Table 5. As we can see, much of the performance difference between IPR and FW can be decreased by the coefficient effects on the social preference parameters. Compared with IPR, FW seems to improve productivity by 2.94, 1.29, 4.75, and

[18] Since the participating rice planters differ slightly each day and some of the variables, such as subjective health and weather conditions, are time-dependent, the endowment difference deviates from zero.

2.91 meters through altruism, guilt aversion, nonenviousness, and voluntary contribution to public goods, respectively.[19] This means that without the social preference effects under FW, the performance gap between IPR and FW could more than triple (Figure 11). In other words, the introduction of monetary incentives under IPR crowded out more than half of the motivation for work under IPR.

The Oaxaca decomposition in Table 5 shows that the increase in performance from switching from FW to IPR is 4.71 meters per 10 minutes or 4.71/25.77 = 18.28%. However, the cost of such a change will be 9.42 PHP per 10 minutes, or 56.4 PHP per hour, because the intensity of the incentive parameter, α, of our IPR contract equals two PHP. A worker engages in rice planting work for around seven hours per day, so the cost of introducing the IPR contract is 394.8 PHP. Since FW is 100 PHP per day, paying an additional 394.8 PHP for a performance increase of 18.28% is unlikely to be socially optimal.

In sum, we obtain three main empirical findings. First, our basic results show positive incentive effects in IPR and moral hazard problems in FW, which are consistent with the standard theoretical implications. Second, under FW, social preferences such as altruism, guilt aversion, and nonenviousness play an important role in stimulating incentives, but introducing monetary incentives crowds out such intrinsic motivations. Third, other non-monetary factors, such as peer effects, significantly change incentives under the FW contract. Considering the significant cost of introducing monetary incentives through IPR, FW seems to be socially optimal.

4.4 Testing Alternative Hypotheses

The Optimal Intensity of the Incentive Condition

The results presented thus far imply that the FW contract can motivate effort via non-monetary incentives. An alternative hypothesis to show the optimality of FW involves a principal’s inability to

[19] The estimated total social preference effect is 11.9 meters per 10 minutes.

accurately measure worker performance. We formulate a simple linearized model of the performance level of each worker observed by the *kabisilya*, η, as follows: $\eta = \delta e + u$, where e is a worker's true effort level, δ is the *kabisilya*'s capacity to measure effort, and u is a measurement error. Then, the rewards schedule offered by the *kabisilya* is $w = F + \phi(\delta e + u)$, where ϕ is the intensity of the incentives. The optimal choice of effort level for each worker is determined by the incentive compatibility constraints: $\delta\phi = C'(e)$, where $C(e)$ is an agent's convex effort cost function. In this setting, the sum of the certainly equivalent for a risk-neutral principal (*kabisilya*) and a risk-averse agent (rice planting worker) is $P(e) - C(e) - \frac{1}{2}rC'(e)^2V$, where $P(e)$ is a principal's expected return function, r is the coefficient of absolute risk aversion, and V is the variance of the effort measurement error, $V = var(u)$. Accordingly, following Milgrom and Roberts (1992), the optimal intensity of incentives becomes:

$$\phi = \frac{P'(e)\delta}{\delta^2 + rC"V}.$$

If $\delta = 0$, the optimal intensity of incentives ϕ becomes zero. This is the case when the FW contract becomes optimal. Hence, we can test the optimality of FW by testing a null hypothesis, $\delta = 0$, in the following estimable equation of a linearized model of the worker performance level observed by the *kabisilya*:

$$\eta_i = \alpha_i + \delta l_{ijt} + X_{ijt}\gamma + u_{ijt}, \quad (3)$$

where we use our *observed* productivity data, l_{ijt}, as a proxy variable for the true effort. This implies that our estimates for δ involve an attenuation bias toward zero. However, even in this case, we can still make an inference about the rejection of the null hypothesis, that is, H_0: $\delta = 0$.

We use each *kabisilya*'s subjective evaluation regarding each worker's performance as the dependent variable, η_i. We estimate the above equation for the FW and IPR samples separately and present the results in Table 6. In all the specifications, the outcomes observed by the *kabisilyas* are positively correlated with measured individual productivity for both FW and IPR. These results seem to be robust even if we include social connection variables to mitigate the evaluation bias that arises from favoritism. Moreover, considering possible errors in measuring the true effort level, l_{ijt}, the estimated δ is likely to involve an attenuation bias. With these results and considerations, we can safely conclude that $\delta > 0$, rejecting FW as the optimal contract when considering monetary incentives only. Hence, for FW to be optimal, it still seems necessary to incorporate non-monetary incentives.

Intertemporal Incentives

Another possibility for FW to be optimal is an intertemporal incentive in contract renewal. If contract renewal probability is a positive function of observed efforts, even under FW, the expected wage level is contingent upon the observed effort level. In this case, the FW contract can be a *de facto* individual piece rate contract. To test this hypothesis, we estimate the following model of contract renewals:

$$EXPER_i = \alpha_i + \rho\eta_{ij} + X_{ij}\gamma + u_{ij},$$

where $EXPER_i$, the years of planting experience under the current *kabisilya*, is a proxy variable for the extent of contract renewals. The set of independent variables is the same as variables shown in Equation (1). If the coefficient on the observed performance level, ρ, is positive and significant, then the model of intertemporal incentives will be supported. Table 7 shows the estimation results. The *kabisilyas*' evaluations are only weakly related to years of planting under the current *kabisilya*. The Blinder–Oaxaca decomposition also indicates that the years of experience under the current *kabisiliya* and the frequency of participation in rice planting during the 2011 dry season are not significantly related to performance under FW (Table 5). These estimation results are inconsistent with the hypothesis of intertemporal incentives.

5. Concluding Remarks

In this paper, we adopted a hybrid experimental method of framed and artefactual field experiments to examine the interplay between economic incentives and social norms. More specifically, to explain the seemingly inefficient FW contract in rice planting, we conducted randomized control trials of three distinct labor contracts as well as two methods of group formation. Three main empirical findings emerge. First, our basic results show the positive incentive effects in IPR, moral hazard problems in FW, and weak free-riding behavior in GPR, which are consistent with the implications of the standard agency theory as well as the empirical findings of Lazear (2000), Foster and Rosenzweign (1994), and Shearer (2004). Second, non-monetary incentives seem to play a significant role under FW, under which an altruistic person with guilt-averse or nonenvious preferences exerts more work efforts in rice planting. However, these roles of social preferences disappear once we introduce the monetary incentives of the IPR contract. These results suggest that extrinsic incentives crowd out intrinsic motivation, a rare real-world finding that confirms evidence from the laboratory experiments reviewed by Bowles (2008). In addition, other non-monetary factors, such as peer effects, significantly enhance work incentives under the FW contract, which is consistent with the findings of Mas and Moretti (2009). Finally, our results are not supportive of the implications of alternative hypotheses—that is, the interlinked contract of labor and credit transactions in mitigating moral hazard problems, the optimality of the FW contract due to large effort measurement errors, and the existence of intertemporal incentives.

Our findings from a hybrid method of framed and artefactual field experiments imply that the introduction of monetary incentives may mitigate moral hazard and free-riding problems but may also generate inefficiency by crowding out intrinsic motivations. Considering the interplay of intrinsic motivations and monetary incentives and the significant costs of introducing IPR, this seemingly perverse traditional contractual arrangements might be socially efficient. The external validity of this finding should be rigorously investigated by implementing carefully designed field experiments in other areas and industries in future studies.

References

Bandiera, Oriana, Iwan Barankay, Imran Rasul (2005) "Social Preferences and the Response to Incentives: Evidence from Personnel Data," *Quarterly Journal of Economics*, 120(3), 917–962.

Bandiera, Oriana, Iwan Barankay, Imran Rasul (2010) "Social Incentives in the Workplace," *Reivew of Economic Studies*, 77, 417–458.

Bandiera, Oriana, Iwan Barankay, Imran Rasul (2013) "Team Incentives: Evidence from a Firm Level Experiment," *Journal of the European Economic Association*, 11(5), 1079-1114.

Bénabou, Roland, and Jean Tirole (2003) "Intrinsic and Extrinsic Motivation," *Review of Economic Studies*, 70(3), 489–520.

Blinder, Alan (1973) "Wage Discrimination: Reduced Form and Structural Estimates," *Journal of Human Resources*, 8, 436–455.

Bowles, Samuel (2008) "Policies Designed for Self-Interested Citizens May Undermine 'The Moral Sentiments': Evidence from Economic Experiments," *Science* 320 (5883), 1605–1609.

Bowles, Samuel and Sandra Polanía-Reyes (2012) "Economic Incentives and Social Preferences: Substitutes or Complements?" *Journal of Economic Literature*, 50(2), 368–425.

Braverman, Avishay and Joseph E. Stiglitz (1982) "Sharecropping and the Interlinking of Agrarian Markets," *American Economic Review*, 72 (4), 695–715.

Camerer, Colin F. (2003), *Behavioral Game Theory*: *Experiments in Strategic Interaction*, New Jersey, Princeton University Press.

Camerer, Colin F., and Ernst Fehr (2004) "Measuring Social Norms and Preferences Using Experimental Games: A Guide for Social Scientists," in Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin F. Camerer, Ernst Fehr, and Herbert Gintis (eds.) *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*, Oxford: Oxford University Press, pp. 55–95.

Cardenas, Juan Camilo and Jeffrey Carpenter (2008) "Behavioural Development Economics: Lessons from Field Labs in the Developing World," *Journal of Development Studies*, 44(3), 311–338.

Carpenter, Jeffrey, Samuel Bowles, Herbert Gintis, Sung-Ha Hwang (2009) "Strong Reciprocity and Team Production: Theory and Evidence," *Journal of Economic Behavior & Organization*, 71, 221–232.

Carpenter, Jeffrey and Erika Seki (2010) "Do Social Preferences Increase Productivity? Field Experimental Evidence from Fishermen in Toyama Bay," *Economic Inquiry*, 49(2), 612–630.

Carpenter, Jeffrey and Tyler Williams (2014) "Peer Monitoring and Microcredit: Field Experimental Evidence from Paraguay," *Oxford Development Studies*, 42(1), 111–135.

David, Cristina C., Violeta G. Cordova, and Keijiro Otsuka (1994), "Technological Change, Land Reform, and Income Distribution in the Philippines," in David, Christina C. and Keijiro Otsuka (eds.) *Modern Rice Technology and Income Distribution in Asia*. Lynne Rienner Publishers, Boulder, Colorado.

Deninger, Klaus, Pedro Olinto, and Miet Maertens (2000) "Redistribution, Investment, and Human Capital Accumulation: The Case of Agrarian Reform in the Philippines," *mimeograph*, World Bank.

Eriksson, Tor and Marie Claire Villeval (2008) "Performance-pay, Sorting and Social Motivation," *Journal of Economic Behavior & Organization*, 68, 412–421.

Estudillo, Jonna P., Yasuyuki Sawada, Mahabub Hossain (2005) "Socioeconomic Changes and the Decline in Poverty: A View from Three Villages in the Philippines, 1985–1997," *Asian Economic Journal*, 19(4), 383–406.

Estudillo, Jonna P., Yasuyuki Sawada, Keijiro Otsuka (2006) "The Green Revolution, Development of Labor Markets, and Poverty Reduction in the Rural Philippines, 1985–2004," *Agricultural Economics*, 35, 399–407.

Estudillo, Jonna P., Yasuyuki Sawada, Keijiro Otsuka (2008) "Poverty and Income Dynamics in Philippine Villages, 1985–2004," *Review of Development Economics*, 12(4), 877–890.

Estudillo, Jonna P., Yasuyuki Sawada, Keijiro Otsuka (2009) "The Changing Determinants of Schooling

Investments: Evidence from Villages in the Philippines, 1985–89 and 2002–04," *Journal of Development Studies*, 45(3), 391–411.

Falk, Armin and Andrea Ichino (2006) "Clean Evidence on Peer Effects," *Journal of Labor Economics*, 24(1), 39–57.

Fehr, Ernst and Klaus M. Schmidt (1999) "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics*, 114(3), 817–868.

Foster, Andrew D. and Mark R. Rosenzweig (1994) "A Test for Moral Hazard in the Labor Market: Contractual Arrangements, Effort, and Health," *Review of Economics and Statistics*, 76(2), 213–227.

Hamilton, Barton H., Jack A. Nickerson, and Hideo Owan. (2003) "Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation," *Journal of Political Economy*, 111(3), 465–497.

Harrison, Glenn W. and John A. List (2004), "Field Experiments," *Journal of Economic Literature* 42(4), 1009–1055.

Hart, Oliver and Bengt Holmstrom (1987) "The Theory Of Contracts," in *Advances in Economic Theory*, Truman Bewley (ed.) Cambridge University Press.

Hayami, Yujiro and Masao Kikuchi (1978) *Anatomy of a Peasant Economy: A Rice Village in the Philippines*, Los Banos, The International Rice Research Institute.

Hayami, Yujiro and Masao Kikuchi (2000) *A Rice Village Saga: Three Decades of Green Revolution in the Philippines*, Lanham/Boulder/New York: Barnes & Noble in association with IRRI.

Itoh, Hideshi (2004) "Moral Hazard and Other-Regarding Preferences," *Japanese Economic Review* 55(1), 18–45.

Kandel, Eugene and Edward P. Lazear (1992) "Peer Pressure and Partnerships," *Journal of Political Economy*, 100, 801–813.

Kerkvliet, Benedict J. Tria. (1990) *Everyday Politics in the Philippines: Class and Status Relations in a Central Luzon Village*. Berkeley: University of California Press.Koszegi, Botond. 2014.

"Behavioral Contract Theory." *Journal of Economic Literature*, 52(4): 1075-1118.

Lazear, Edward P. (1995) *Personnel Economics*. MIT Press: Cambridge.

Lazear, Edward P. (2000) "Performance Pay and Productivity," *American Economic Review*, 90(5), 1346–1361.

Ledyard, John (1995) "Public Goods: A Survey of Experimental Research," in *Handbook of Experimental Economics*, J. Kagel and A. Roth (eds.), pp. 111–194. Princeton, NJ: Princeton University Press.

Levitt, Steven D. and John A. List (2009), "Field Experiments in Economics: The Past, the Present, and the Future," *European Economic Review*, 53, 1–18.

Levitt, Steven D. and John A. List (2007) "What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?" *Journal of Economic Perspectives* 21(2), pp. 153-174

List, John and Imran Rasul (2011), "Field Experiments in Labor Economics," in *Handbook of Labor Economics* Vol.

Mas, Alexandre and Enrico Moretti (2009) "Peers at Work," *American Economic Review*, 99, 112–145.

Munro, Alistair, Arjan Verschoor, and Amaresh Dubey (2013) "Does Working with Spouses Make Teams more Productive? A Field Experiment in India using NREGA," *Economic Letters*, 118, 506–508.

Oaxaca, Ronald (1973), "Male–Female Wage Differentials in Urban Labor Markets," *International Economic Review*, 14, 693–709.

Otsuka, Keijiro (1991) "Determinants and Consequences of Land Reform Implementation in the Philippines," *Journal of Development Economics*, 35(2), 339–355.

Otsuka, Keijiro, Hiroyuki Chuma, and Yujiro Hayami (1993) "Permanent Labour and Land Tenancy Contracts in Agrarian Economies: An Integrated Analysis," *Economica*, 60, 57–77.

Radner, Roy (1986) "Repeated Partnership Games with Imperfect Monitoring and No Discounting," *Review of Economic Studies*, 53(1), 43–57.

Roth, Alvin E., Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir (1991) "Bargaining and

Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study," *American Economic Review*, 81(5), 1068–1095.

Roumasset, James and Sang-Hyop Lee (2007) "Labor: Decisions, Contracts and Organization," Robert Evenson and Prabhu Pingali (eds.), *Handbook of Agricultural Economics*, North Holland, Elsevier, chapter 52, pp. 2705–2740.

Sausgruber, Rupert (2009) "A Note on Peer Effects between Teams" *Experimental Economics*, 12, 193–201.

Schechter, Laura (2007), "Traditional Trust Measurement and the Risk Confound: An Experiment in Rural Paraguay," *Journal of Economic Behavior & Organization*, 62, 272–292.

Shapiro, Carl and Joseph E. Stiglitz (1984), "Equilibrium Unemployment as a Worker Discipline Device," *American Economic Review*, 74(3), 433-44.

Shearer, Bruce (2004) "Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment," *Review of Economic Studies*, 71 (2), 513–534.

Umekage, Osamu, Koichiro Inoue, and Tadayo Watabe (1966), "Studies on the Simplification of Transplanting Methods of Rice Plants I: Growth Habits of the Plants from Which Seedlings are Horizontally Transplanted" (in Japanese), *Kyoto Prefectural University Academic Report on Agriculture* (Kyoto Furitsu Daigaku Gakujyutsu Houkoku Nougaku), 18, 70–77.

Picture 1: Rice Planting

Figure 1: Structure of a planting group

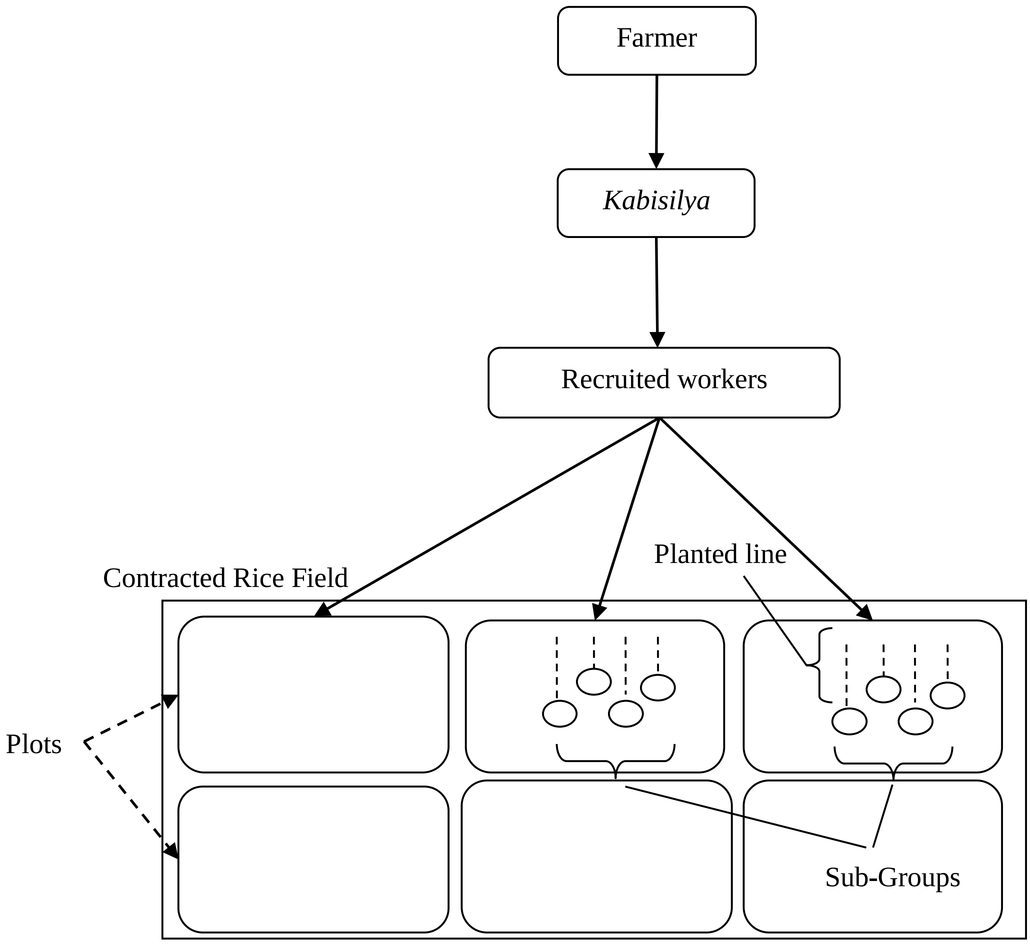

Figure 2: Timeline of experiments

Date (Jan, 2011)		11		12		13		14		15		17		18		19		20		21		22		24		
Time		AM	PM	AM	PM	AM	PM	AM	PM	AM	PM	AM	PM	AM	PM	AM	PM	AM	PM	AM	PM	AM	PM	AM		
Group Selection		Random-selection								Self-selection										Self-selection						
G village	Round	1	2	3		4	5	6	7	8	9	10			11	12	13	Artefactual field experiment				14	15		Sub-total	Total
	FW	x				x			x			x				x							x		6	15
	GPR			x				x		x					x										4	
	IPR		x				x				x						x					x			5	

Group Selection		Random-selection												Self-selection						Self-selection						
M village	Round	1		2	3		4				5	6	7		8	9	10			Artefactual field experiment		11	12	13	Sub-total	Total
	FW	x					x					x				x							x		5	13
	GPR				x								x		x							x			4	
	IPR			x							x						x							x	4	

Note: x indicates the experiment we conducted. The first two rounds were dropped from our analysis because the fixed wage component is set at 20 PHP in these rounds.

Table 1: Descriptive statistics

Panel A:Productivity	Obs	Mean	Std. Dev.	Min	Max
Productivity (stack for 1–10, 10–20, 20–30 min)					
Fixed wage	732	26.07632	9.226555	1.896	69.144
Individual piece rate	522	29.96725	9.370332	2.01915	65.388
Group piece rate	453	29.74732	9.205822	6.4725	69.8
Panel B: Field Experiment Condition	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
Average years of acquaintance with group members	569	10.63813	9.437524	0	42.4
Share of group members attending the same church	569	0.3928237	0.3917962	0	1
Self-selection dummy	569	0.4604569	0.4988725	0	1
Plot condition (base category: bad)					
Usual	569	0.5465729	0.4982643	0	1
Good	569	0.3585237	0.4799889	0	1
Weather (base category: clear and sunny)					
Sunny	569	0.3216169	0.4675078	0	1
Cloudy	569	0.2565905	0.4371357	0	1
Wind force (base category: no wind)					
Breeze	569	0.4973638	0.500433	0	1
Weak	569	0.059754	0.237239	0	1
Strong	569	0.1757469	0.3809396	0	1
Missing wind force	569	0.0228471	0.1495474	0	1
Temperature	569	22.93146	12.99501	0	38
Missing temperature	569	0.2337434	0.4235832	0	1
Health condition (base category: good)					
Same as usual	548	0.4744526	0.4998031	0	1
Bad	548	0.0145985	0.1200488	0	1
Experimenter dummy	569	0.4253076	0.4948246	0	1
G village dummy	569	0.5202109	0.5000309	0	1
Plant consecutively	569	0.7486819	0.4341527	0	1
PM session dummy	569	0.4956063	0.5004206	0	1
Interval (base category: 0–10 min)					
10–20 min	569	0.340949	0.4744454	0	1
20–30 min	569	0.3268893	0.4694891	0	1
Group size	569	5.369069	0.8417951	3	7
Participation number	569	3.620387	2.228182	1	10
Panel C: Individual Characteristics	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
# of participations/total field experiment sessions	115	0.4138036	0.2122222	0.0769231	0.8181818
Occupation type (base category: non-agricultural)					
Farmer	110	0.2	0.4018307	0	1
Daily agricultural worker (landless)	110	0.4181818	0.4955179	0	1
Porcientuhan[+]	110	0.2909091	0.4562603	0	1
Other	110	0.0090909	0.0953463	0	1
Age	111	35.59459	15.89304	13	80
Age (squared)	111	1517.288	1257.784	169	6400
Sex (male = 1)	111	0.3423423	0.4766454	0	1
Years of schooling	109	7.036697	2.613805	1	13
Years of experience under current *kabisiliya*	104	4.076923	5.354324	0	30
Advanced payment (PHP)	102	625.4902	617.1936	0	3000
Relationship with *kabisiliya*					
Relative or family of *kabisiliya*	104	0.3846154	0.4888602	0	1
Friend of *kabisiliya*	104	0.2788462	0.4506033	0	1
Neighbor of *kabisiliya*	104	0.1826923	0.3882853	0	1

Note) *Porcientuhan* refers to a landless worker who is employed to work on the farm for a payment of 10 percent of the gross value of output.

Figure 3: Cumulative distribution function of individual productivity

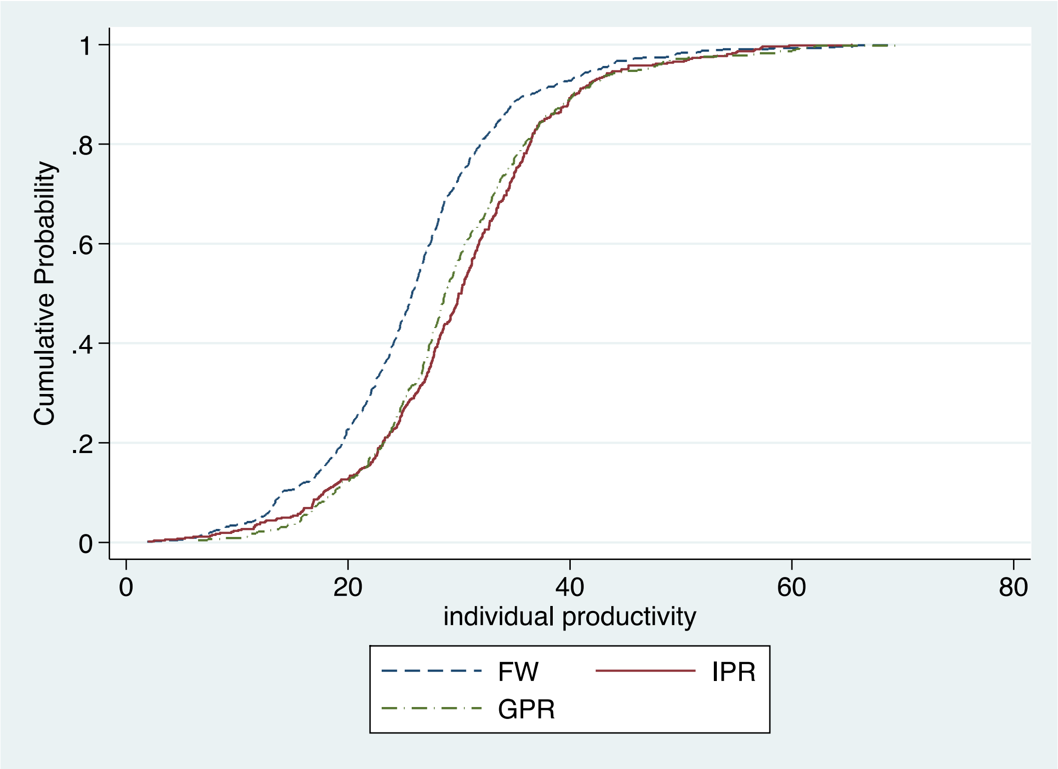

Figure 4: Individual and average productivity of peers

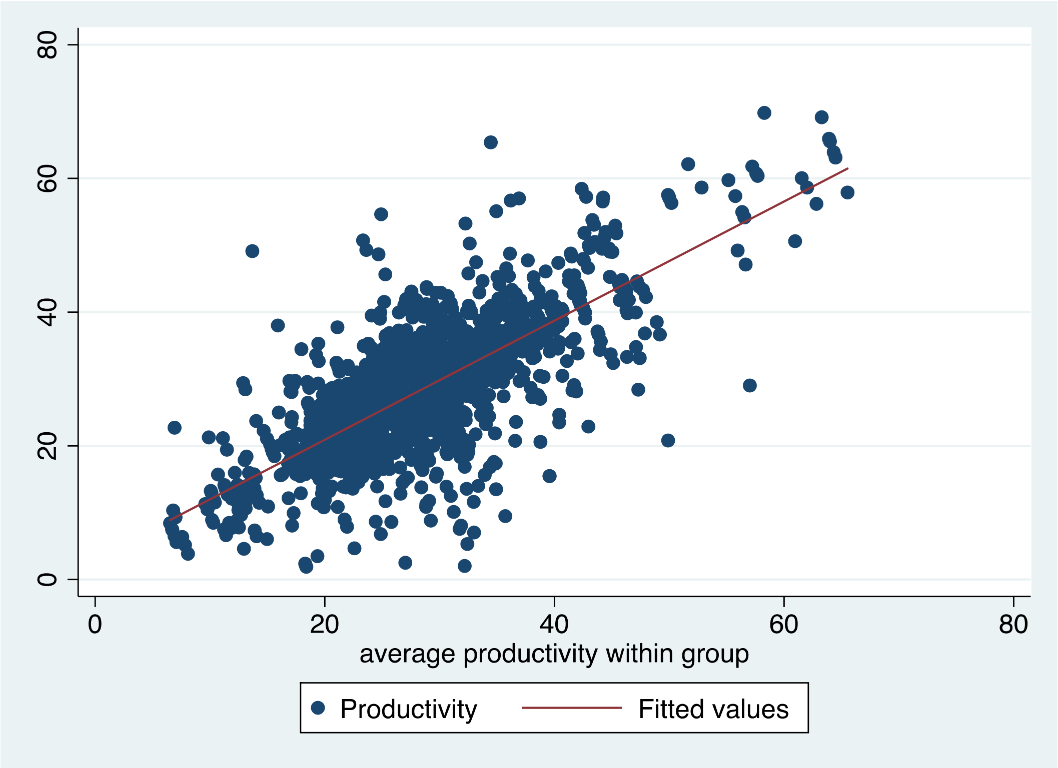

Table 2: Descriptive statistics of artefactual field experiments

Panel A: Dictator and Ultimatum Games	Obs	Mean	Std. Dev.	Min	Max
Dictator game (sending amount)	104	27.21154	18.45786	0	100
Ultimatum game (sending amount)	103	35.04854	15.77251	10	90
Ultimatum game (responder's minimum acceptance level)	104	2.865385	2.755604	0	10

Panel B: Public Goods Game	Obs	Mean	Std. Dev.	Min	Max
Contribution in 1st round	104	33.26923	20.64223	0	100
Contribution in 2nd round	104	31.44231	21.8298	0	100
Contribution in 3rd round	104	35.19231	27.51807	0	100
Average contribution	104	33.30128	19.21038	10	86.66666
Monitoring in 1st round	104	0.4615385	0.5009327	0	1
Monitoring in 2nd round	104	0.3461538	0.4780468	0	1
Monitoring in 3rd round	104	0.3076923	0.4637735	0	1
# of disapproval messages in 1st round	104	0.1346154	0.3955503	0	2
# of disapproval messages in 2nd round	104	0.2115385	0.5689895	0	3

Panel C: Risk Game	Obs	Mean	Std. Dev.	Min	Max
Invested amount	104	37.69231	22.4357	10	100

Figure 5: Proposer's offer in the dictator game

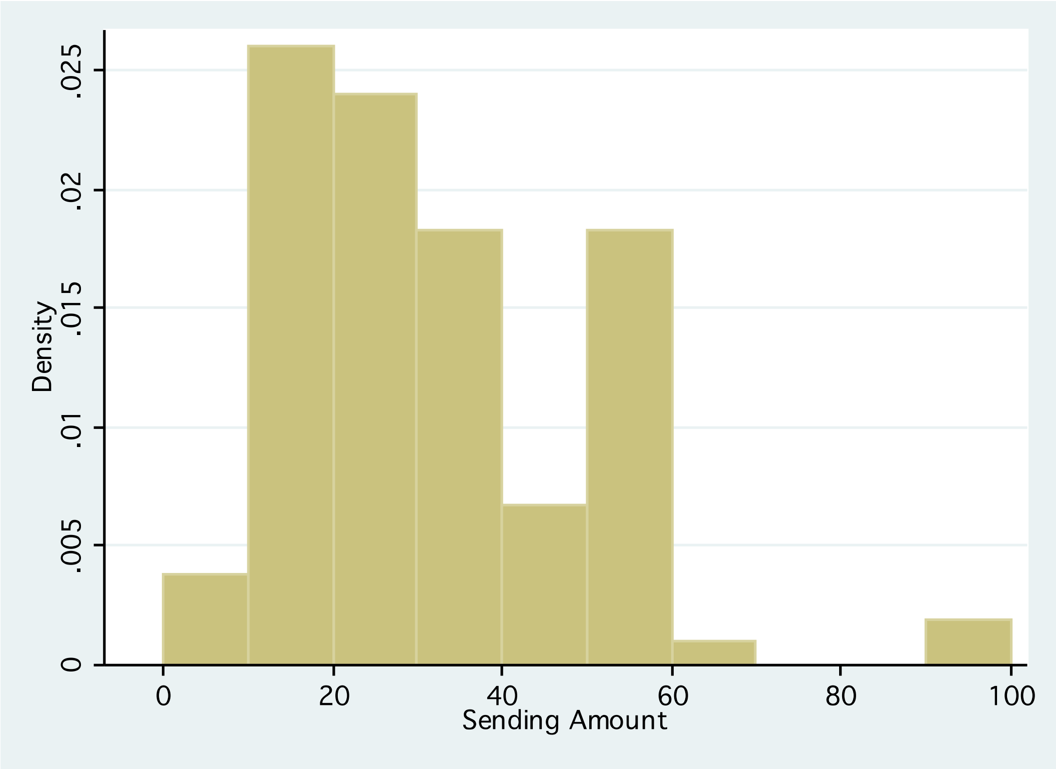

Figure 6: Proposer's offer in the ultimatum game

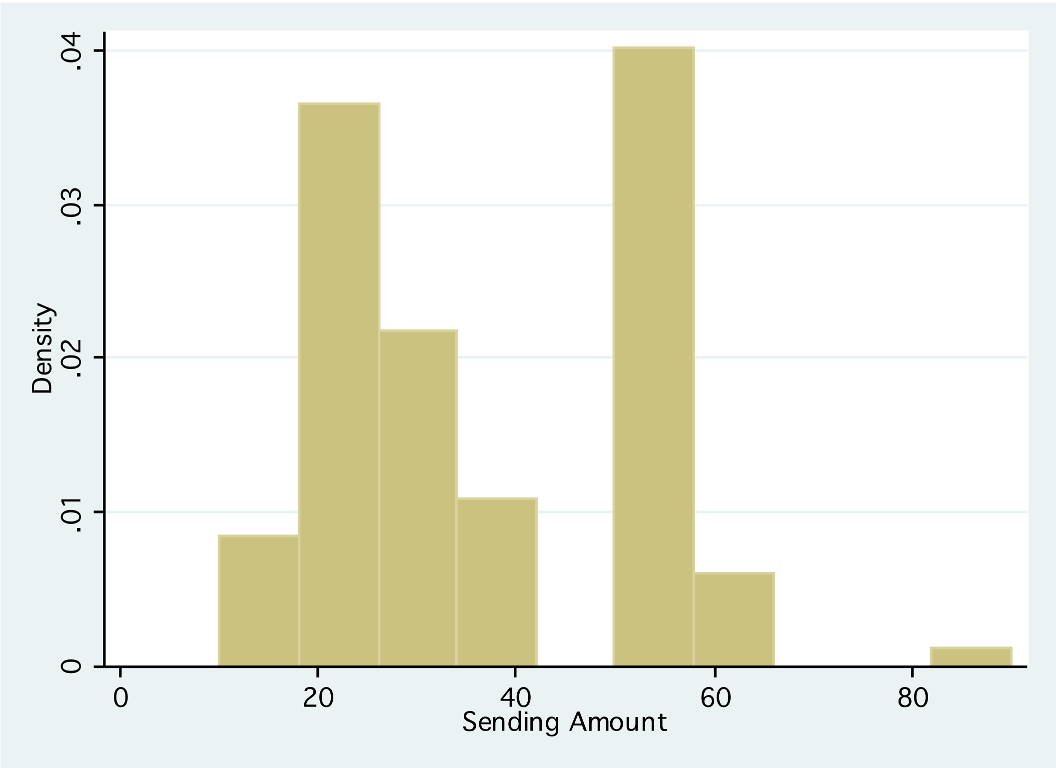

Figure 7: Responder's acceptance in the ultimatum game

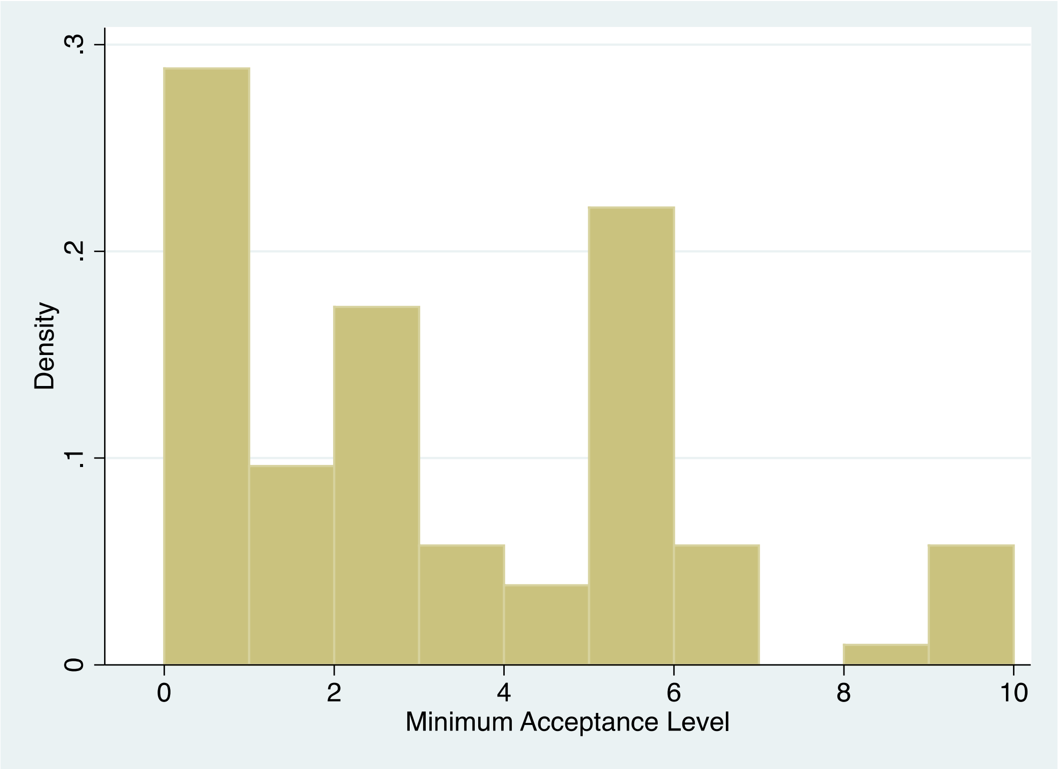

Figure 8: Mean contribution level in the public goods game

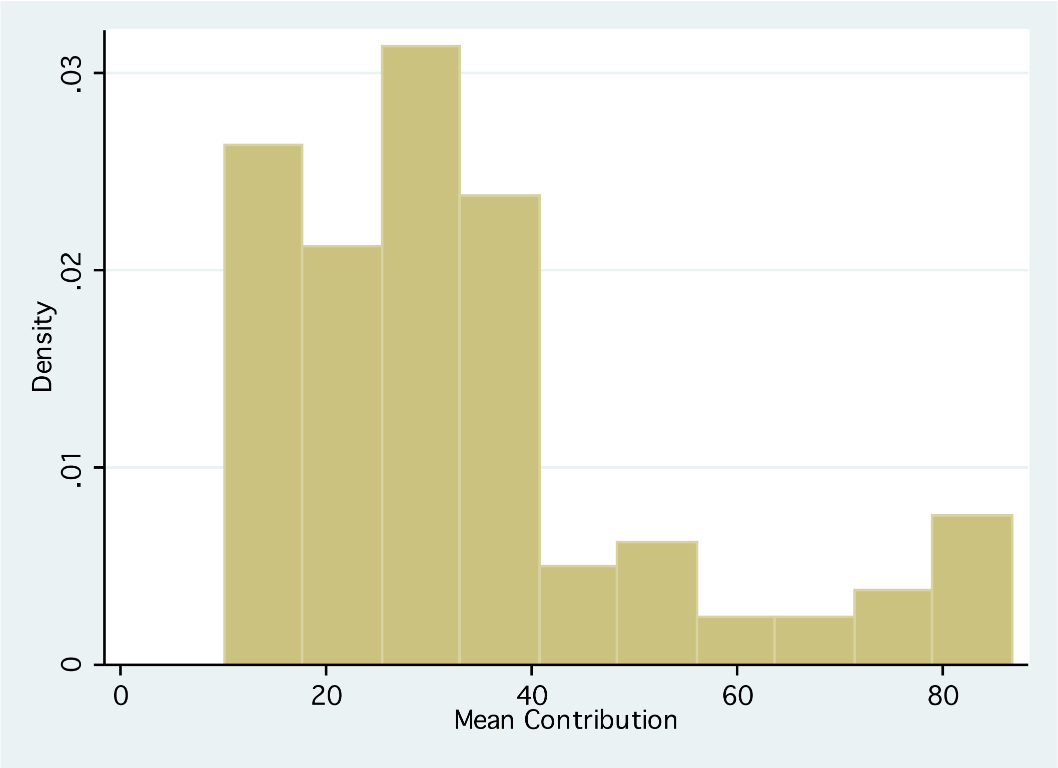

Figure 9: Investment level in the risk game

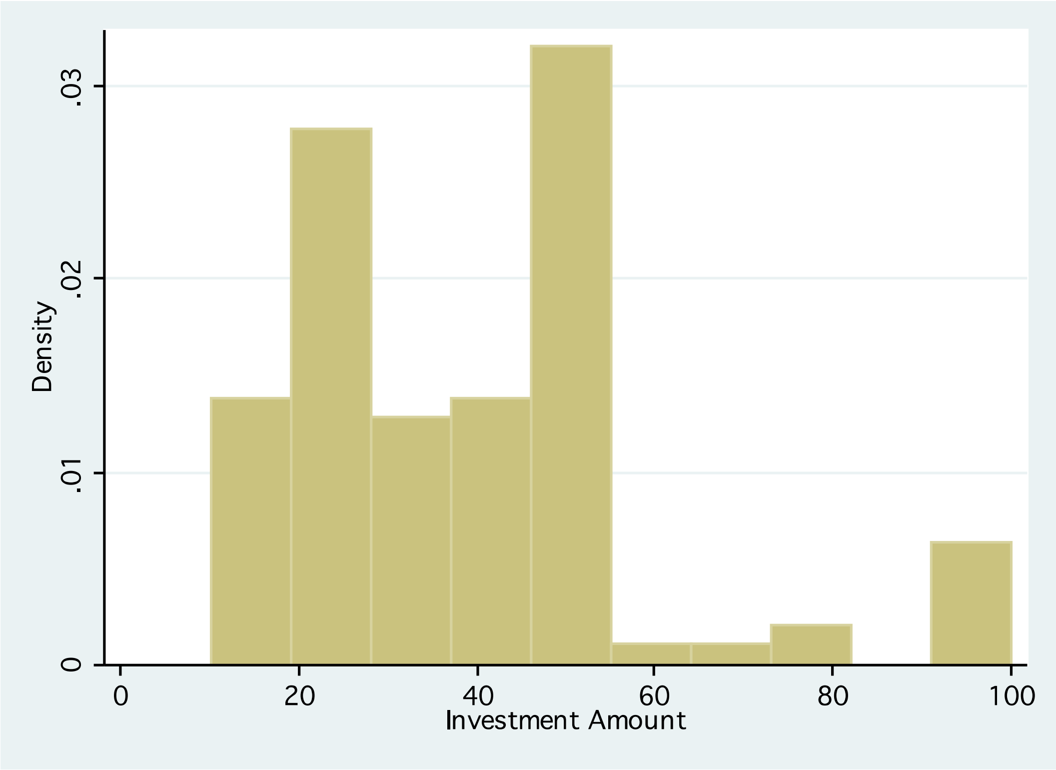

Figure 10: Distribution of estimated permanent ability

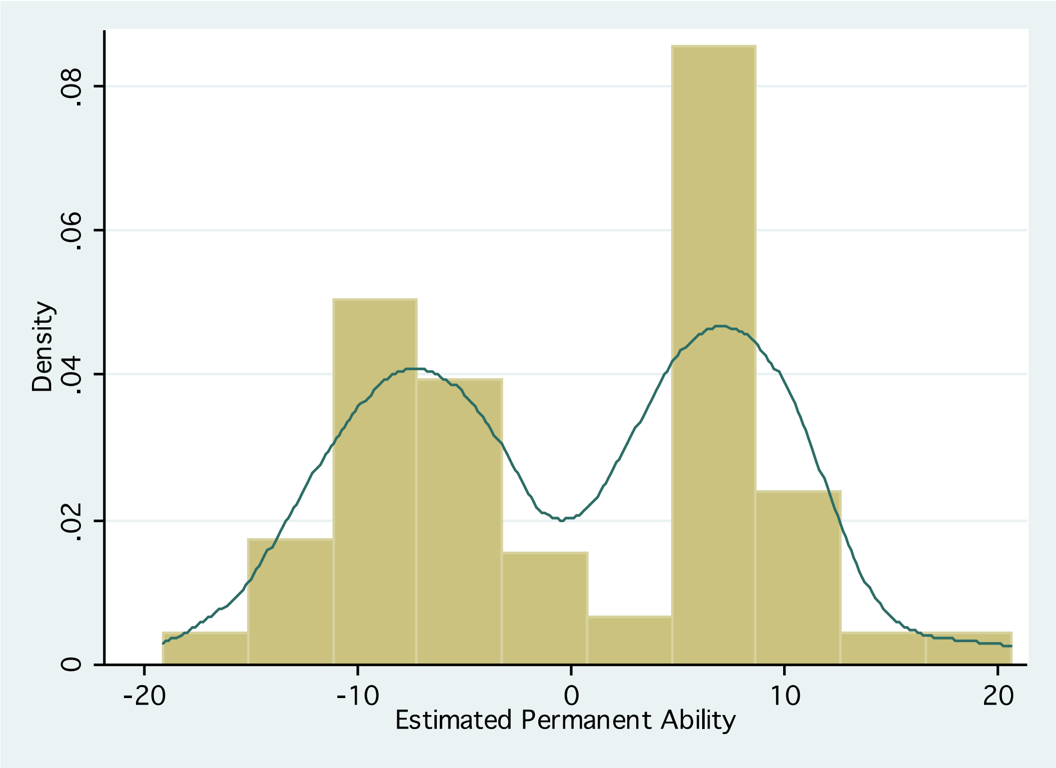

Figure 11: Oaxaca Decomposition of the Performance Gap between IPR and FW Contracts

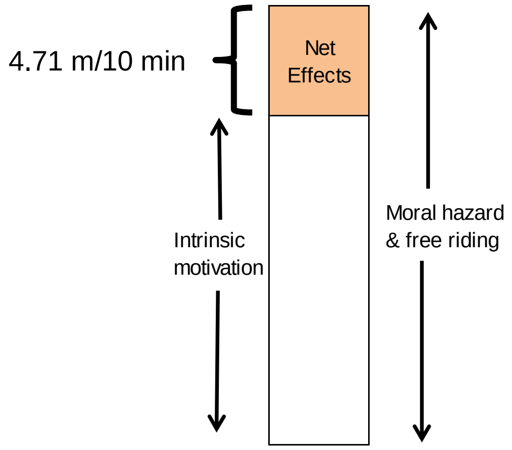

Table 3: Estimation results of rice planters' work performance regression
(The dependent variable is the length of planted line per 10 minutes)

	(1)	(2)	(3)	(4)
Fixed wage dummy	-3.608***	-3.651***	-3.665***	-3.745***
	(1.022)	(1.074)	(0.870)	(0.868)
Group piece rate dummy	-0.808	-1.266	-0.963	-0.870
	(1.016)	(1.030)	(0.868)	(0.867)
Dictator game		0.0531	0.0796***	0.0822***
		(0.0328)	(0.0222)	(0.0229)
Guilt aversion (amount sent in the ultimatum game		0.0905***	0.0918***	0.0775***
– amount sent in the dictator game)		(0.0268)	(0.0212)	(0.0231)
Nonenviousness (10 coins – minimum acceptance		0.465***	0.402***	0.463***
level in the ultimatum game)		(0.172)	(0.108)	(0.111)
Average cooperation in PGG		0.0257	0.00751	-0.00225
		(0.0266)	(0.0203)	(0.0214)
Risk game		-0.0277	-0.0406***	-0.0389**
		(0.0203)	(0.0157)	(0.0164)
Average ability of the other group members			0.749***	0.742***
			(0.184)	(0.184)
Self-selection dummy			-0.692	-0.869
			(1.026)	(1.033)
Years of experience under the current *kabisiliya*			0.0732	0.0534
			(0.0554)	(0.0570)
Relative or family of *kabisiliya*				-2.403***
				(0.924)
Friend of *kabisiliya*				-2.031**
				(0.918)
Neighbor of *kabisiliya*				-0.141
				(1.122)
Average years of acquaintance with group members				0.0597*
				(0.0335)
Share of group members attending the same church				0.888
				(0.723)
Advanced payment (PHP)			0.000707	0.000251
			(0.000451)	(0.000468)
# of participations/total field experiment sessions	8.853***	8.611***	5.903***	6.744***
	(2.949)	(2.923)	(2.119)	(2.170)
Field Experiment Conditions	YES	YES	YES	YES
Individual Characteristics	YES	YES	YES	YES
Observations	1584	1488	1371	1371
R-squared	0.186	0.206	0.230	0.239

Note) Clustered standard errors are reported in parentheses in (1) and (2) and the bootstrapped SEs in (3) and (4). We include the following variables. Field Experiment Conditions: plot condition dummies, weather dummies, wind force dummies and its missing dummy, temperature and its missing dummy, health condition dummies, G village dummy, experimenter dummy, consecutive planting dummy, PM dummy, interval dummies, group size dummies, and participation number dummies. Individual Characteristics: occupation type dummies, age, age squared, sex, and years of schooling.

Table 4: Estimation results of rice planters' work performance regression by contract type
(The dependent variable is the length of planted line per 10 minutes)

	(1)	(2)	(3)	(4)	(5)	(6)
Type of contract	FW	IPR	GPR	FW	IPR	GPR
Dictator game	0.0925***	-0.0346	0.0283	0.119***	0.00532	0.0448
	(0.0302)	(0.039)	(0.037)	(0.0333)	(0.0492)	(0.0446)
Guilt aversion (amount sent in the ultimatum	0.106***	-0.00152	0.0854*	0.0978***	-0.0575	0.108**
– amount sent in the dictator game)	(0.0310)	(0.0407)	(0.0369)	(0.0349)	(0.0514)	(0.0478)
Nonenviousness (10 coins – minimum acceptance	0.614***	-0.153	0.403**	0.585***	-0.146	0.443**
level in the ultimatum game)	(0.184)	(0.170)	(0.199)	(0.190)	(0.183)	(0.219)
Average cooperation in PGG	0.0333	-0.0418	0.0347	0.0204	-0.0773*	0.0298
	(0.0268)	(0.0382)	(0.0348)	(0.0305)	(0.0450)	(0.0419)
Risk game	-0.0598	0.0399	-0.0503*	-0.0511**	-0.00735	0.0639*
	(0.0227)	(0.0280)	(0.0262)	(0.0238)	(0.0333)	(0.0322)
Average ability of the other group members	0.730***	0.168	0.464	0.666**	0.290	0.668
	(0.266)	(0.400)	(0.379)	(0.277)	(0.442)	(0.410)
Self-selection dummy				-11.16	-4.035	-
				(9.177)	(3.052)	(2.041)
Years of experience under the current *kabisiliya*				0.0745	0.237**	-0.126
				(0.0859)	(0.109)	(0.110)
Relative or family of *kabisiliya*				-2.288*	-0.887	-
				(1.360)	(1.804)	(1.734)
Friend of *kabisiliya*				-1.422	0.851	-
				(1.329)	(1.861)	(1.753)
Neighbor of *kabisiliya*				1.270	0.243	-4.127*
				(1.734)	(2.037)	(2.293)
Average years of acquaintance with group				0.0997**	-0.0447	0.0303
				(0.0519)	(0.0679)	(0.0605)
Share of group members attending the same				1.491	2.839*	0.148
				(1.075)	(1.487)	(1.466)
Advanced payment (PHP)				-3.59×10^{-05}	0.00025	0.00032
				(0.0007)	(0.0008)	(0.0010)
# of participations/total field experiment sessions	4.550	16.39**	3.264	3.310	16.63**	0.890
	(3.559)	(4.107)	(3.655)	(3.646)	(4.607)	(4.455)
Field Experiment Conditions	YES	YES	YES	YES	YES	YES
Individual Characteristics	YES	YES	YES	YES	YES	YES
F stat on social preference	7.84***	0.98	1.18	11.25***	1.50	1.23
Observations	627	456	405	582	426	363
R-squared	0.268	0.312	0.348	0.314	0.296	0.374

Note) Bootstrapped standard errors are reported in parentheses. The F-statistics on social preferences are calculated based on clustered standard errors. We include the following variables. Field Experiment Conditions: plot condition dummies, weather dummies, wind force dummies and its missing dummy, temperature and its missing dummy, health condition dummies, G village dummy, experimenter dummy, consecutive planting dummy, PM dummy, interval dummies, group size dummies, and participation number dummies. Individual Characteristics: occupation type dummies, age, age squared, sex, and years of schooling.

Table 5: Blinder–Oaxaca Decomposition of the Performance Difference between the Fixed Wage and the Individual Piece Rate

VARIABLES	(1) Differential	(2) Endowments	(3) Coefficients	(4) Interaction
Prediction (Fixed wage)	25.77***			
	(0.542)			
Prediction (Individual piece rate)	30.48***			
	(0.578)			
Difference	-4.711***			
	(0.711)			
Dictator game		0.00802	2.944**	0.171
		(0.0780)	(1.300)	(0.213)
Guilt aversion (amount sent in the ultimatum game – amount sent in the dictator game)		-0.0142	1.291**	0.0383
		(0.102)	(0.561)	(0.275)
Nonenviousness (10 coins – minimum acceptance level in ultimatum game)		-0.0929	4.754***	0.466**
		(0.122)	(1.374)	(0.232)
Average cooperation in PGG		-0.292	2.906*	0.369
		(0.211)	(1.605)	(0.249)
Risk game		-0.0289	-1.494	-0.172
		(0.104)	(1.355)	(0.170)
Average ability of the other group members		0.548	-0.882	0.709
		(0.801)	(1.228)	(0.965)
Self-selection dummy		-0.0797	-3.310	-0.141
		(0.142)	(4.157)	(0.295)
Years of experience under the current *kabisiliya*		-0.0879	-0.745	0.0604
		(0.128)	(0.494)	(0.0933)
Relative or family of *kabisiliya*		-0.0375	-0.641	-0.0592
		(0.0684)	(0.840)	(0.0964)
Friend of *kabisiliya*		-0.0100	-0.624	0.0267
		(0.0372)	(0.541)	(0.0867)
Neighbor of *kabisiliya*		-0.00838	0.152	-0.0354
		(0.0587)	(0.335)	(0.0818)
Average years of acquaintance with group members		0.0242	1.692*	-0.0783
		(0.0521)	(0.937)	(0.134)
Share of group members attending the same church		0.143	-0.531	-0.0680
		(0.115)	(0.635)	(0.0927)
Advanced payment (PHP)		-0.0145	-0.220	0.0165
		(0.0472)	(0.679)	(0.0532)
# of participations/total field experiment sessions		-0.496	-7.344***	0.398
		(0.321)	(2.405)	(0.268)
Total		1.700	-7.216	0.804
		(1.330)	(6.196)	(6.260)
Field Experiment Conditions	YES	YES	YES	YES
Individual Characteristics	YES	YES	YES	YES
Observations	1008	1008	1008	1008

Note) Bootstrapped standard errors are reported in parentheses. We include the following variables. Field Experiment Conditions: plot condition dummies, weather dummies, wind force dummies and its missing dummy, temperature and its missing dummy, health condition dummies, G village dummy, experimenter dummy, consecutive planting dummy, PM dummy, interval dummies, group size dummies, and participation number dummies. Individual Characteristics: occupation type dummies, age, age squared, sex, and years of schooling.

Table 6: Test of the optimal intensity of incentives

(The dependent variable is the *kabisiliya*'s subjective evaluation of each planting worker's performance)

	(1) FW	(2) FW	(3) FW	(4) IPR	(5) IPR	(6) IPR
Productivity	0.0437***	0.0350**	0.0357**	0.0769***	0.0451***	0.0457***
	(0.0163)	(0.0149)	(0.0147)	(0.0170)	(0.0155)	(0.0150)
Relative or family of *kabisiliya*			0.462			1.136
			(0.769)			(1.055)
Friend of *kabisiliya*			0.456			0.766
			(0.850)			(1.099)
Neighbor of *kabisiliya*			0.304			1.000
			(0.864)			(0.933)
# of participations/total field experiment sessions		1.544	1.414		1.328	1.090
		(2.043)	(2.069)		(2.134)	(2.255)
Field Experiment Conditions	NO	YES	YES	NO	YES	YES
Individual Characteristics	NO	YES	YES	NO	YES	YES
Observations	678	627	627	492	459	456
R-squared	0.023	0.391	0.394	0.066	0.514	0.526

Note) Clustered standard errors are in parentheses. We include the following variables. Field Experiment Conditions: plot condition dummies, weather dummies, wind force dummies and its missing dummy, temperature and its missing dummy, health condition dummies, G village dummy, experimenter dummy, consecutive planting dummy, PM dummy, interval dummies, group size dummies, and participation number dummies. Individual Characteristics: occupation type dummies, age, age squared, sex, and years of schooling.

Table 7: Test of intertemporal incentives

(The dependent variable is years of work experience under the current *kabisiliya*)

VARIABLES	(1)	(2)	(3)
Kabisiliya's subjective evaluation	0.308*	0.305	0.310
	(0.166)	(0.207)	(0.215)
Relative or family of *kabisiliya*			-1.294
			(1.960)
Friend of *kabisiliya*			0.774
			(2.259)
Neighbor of *kabisiliya*			-2.211
			(1.781)
Individual characteristics	NO	YES	YES
Observations	103	100	100
R-squared	0.024	0.262	0.300

Note) Clustered errors are in parentheses. We include the following variables. Individual Characteristics: # of participations/total field experiment sessions, occupation type dummies, age, age squared, sex, years of schooling, and G village dummy.