Gender, self-selection and moral hazard in teams

Eberhard Feess[a], Chiara Nardi[a,*]

[a] *Frankfurt School of Finance & Management, Frankfurt, Germany*

This draft: March 24, 2017

Abstract

Motivated by previous research on incentive design in organizations, we conduct a laboratory experiment that investigates the role of specific contract features in the effort choice, and examines if and how these contract features interact with gender. We compare an individual payment scheme to a team payment scheme in which partners equally share the produced output and bear effort costs individually. We distinguish between random assignment and self-selection to payment schemes. Our analysis reveals interesting results. First, females self-select into team payment more often than males. Second, females who are randomly assigned to team payment are more likely to choose the first-best (cooperative) effort level than males. Third, males (but not females) who prefer to be rewarded with team payment are more likely to choose the cooperative effort level than those who are randomly assigned to it. Finally, both the payment choice and the effort choice are driven by beliefs on the partner's behavior, irrespectively of gender.

Keywords: Payment schemes, Self-selection, Gender, Experiment
JEL Classification: C91, J16, J24, J33, M5

*Corresponding author. Frankfurt School of Finance & Management, Sonnemannstrasse 9-11, Frankfurt am Main (Germany). Tel.: 0049 69 154008 705; fax: 0049 69 154008 4705. *Email*: c.nardi@fs.de

1 Introduction

Moral hazard in teams arises because effort costs are borne individually, while output is shared among team members (Alchian and Demsetz, 1972; Newhouse, 1973; Holmström, 1982). The degree of moral hazard in teams is known to depend on multiple factors such as the institutional setting (Gächter, 2007), team heterogeneity (Hamilton et al., 2003), fairness preferences and other personality traits (Gächter and Thöni, 2005; Glew, 2009; Bradley et al., 2013), and the expectations on the behavior of other participants (Keser and van Winden, 2000; Frey and Meier, 2004; Ball and Eckel, 1998; Fischbacher and Gächter, 2010). Group composition has been studied from many perspectives (for reviews, see Prendergast, 1999 and Mathieu et al., 2014), but little is known about the impact of self-selection on efficiency in teams (a notable exception is Kuhn and Villeval, 2015). As self-selection to teams may attract selfish participants hoping to free-ride on their partner's effort as well as subjects with a high sense for cooperation, it is not obvious whether self-selection leads to higher or lower moral hazard in teams compared to random assignment.[1] In a chosen effort-experiment, we find that the impact of self-selection on the degree of moral hazard in teams differs largely between males and females. While females are more cooperative than males when randomly assigned to teams, the degree of moral hazard of males shrinks largely when they self-select to teams.

In our experiment, subjects decide between individual compensation and revenue-sharing in teams. To avoid potentially confounding effects of differences in capabilities and intrinsic motivation, we adopt chosen effort instead of real effort. Subjects see tables with deterministic effort costs and revenues, from which they can easily calculate the payoff-maximizing effort levels. As the production function is additively separable, we neglect synergies in the spirit of Holmström (1982)'s team production. Thus, the cooperative effort that maximizes the team's joint payoff is the same as in the individual setting, but the payoff-maximizing effort is lower in teams due to revenue-sharing. Neglecting synergies avoids that the payoff maximizing own effort depends on the belief on the partner's effort, which would hamper a clear-cut interpretation of the experimental findings.

[1] Kosfeld and von Siemens (2009) derive a separating equilibrium in which workers with high and low exposure to opportunistic behavior in teams self-select to different firms, so that firms employing more cooperative workers earn higher profits.

Considering all subjects together, we observe no significant differences between random assignment and self-selection, but there are large gender differences for the impact of self-selection on effort levels. Our first result is that females are significantly more likely to self-select into teams compared to choosing individual payment, while there is no difference for males. For both genders, choosing teams is largely driven by the expectation on the partner's effort level; the higher the expected degree of cooperation, the higher is the likelihood of self-selecting into teams. Second, when randomly assigned to teams, females choose the cooperative effort level significantly more often than males. Third, however, the gender impact on effort levels in teams is reversed when considering self-selection instead of random assignment: While less than 23% of males choose the cooperative effort level when randomly assiged to teams, this percentage increases to 59% with self-selection. By contrast, there is no positive impact of self-selection for females. Thus, while males show a higher degree of moral hazard than females with random assignment, they cooperate more when it is their own decision to share the revenue with their partner.

Our research is complementary to Kuhn and Villeval (2015), who seems the only other paper considering the impact of gender on self-selection and effort in teams. Conversely to our approach, they apply a real effort task consisting of decoding numbers into letters, so that self-selection is also influenced by capabilities. As low performers are more likely to benefit from partners, team performance is reduced by adverse selection. This adverse selection effect is larger for males who have a stronger tendency to overestimate their capabilities relative to others. Furthermore, Kuhn and Villeval (2015) allow for synergies from team production,[2], so that the socially optimal efforts depend on capabilities. Instead, we trigger self-selection into teams by offering a fixed fee, while optimal efforts are the same in all treatments.

In line with empirical results (Encinosa III et al., 2007; Pizzini, 2010), laboratory experiments by Chao and Croson (2013) and Bäker and Pull (2016) find that teams outperform individual pay if and only if synergies are sufficiently strong. Keser and Montmarquette (2011) show that self-selection into teams increases payoffs compared to random assignment; a

[2]Specifically, they multiply the piece rate for each unit of output by a factor $r > 1$.

result we obtain only for male participants.[3] By contrast to these papers, there are neither adverse selection nor synergies in our setting, and our additively separable production function allows us to attribute any effect of self-selection exclusively to the impact on the degree of moral hazard.

The real-effort experiment by Bäker and Mertins (2013) extends the choice between individual and team payment to stochastic output, so that team payment allows for risk-sharing. Lower effort in teams is fully attributable to adverse selection effects, while there is no moral hazard effect, which is often the case in real-effort experiments when participants face no opportunity costs.[4] Participants with higher risk aversion tend to self-select into team payments in order to reduce the risk. By contrast, the payoff in our experiment is uncertain from each participant's point of view only in teams, so that the likelihood of choosing the team decreases in risk aversion (only for males; however). Other papers consider either the choice between fixed payments and several kinds of variable payments (Cadsby et al., 2007; Eriksson and Villeval, 2008; Dohmen and Falk, 2011) or the impact of gender composition in teams without extending to self-selection (Ivanova-Stenzel and Kübler, 2011; Apesteguia et al., 2012; Delfgaauw et al., 2013).

In addition to experiments on team performance, VCM-experiments are related due to comparable free-rider incentives, but gender effects are inconclusive (see the surveys by Eckel and Grossman, 2008 and Croson and Gneezy, 2009. As to self-selection, Nosenzo and Tufano (2015) find that the threat of exiting a VCM-game increases cooperation, but find no positive impact of self-selecting into participating in the game.

The remainder of the paper is organized a follows. Section 2 describes the experimental design and procedures. Section 3 presents the results. Section 4 summarizes the main findings of the study and offers concluding remarks.

2 The experiment

The experiment consisted of five parts. Participants were informed at the beginning that there would be five parts, but immediately received only the

[3] Vyrastekova et al. (2012) find that participants exhibiting high levels of trust and reciprocity are more likely to self-select into team payment.

[4] Performances in team payments tend to go down when opportunity costs are implemented, for instance by the possibility of browsing the internet (Mohnen et al., 2008; Corgnet et al., 2015).

instructions for Part 1.

In all parts monetary amounts were expressed in ECU (Experimental Currency Units), with a conversion rate of 1 ECU = 0.10 Euro. In order to prompt participants to truthfully report their choices in each and every part of the experiment, we used the random incentive system.[5] At the end of the experiment only one of the parts was randomly chosen for payout, where each part had the same probability of being selected. Participants knew about this procedure from the very beginning and were informed about the selected part and the associated earnings at the end of the session.

In what follows, we will first describe the experimental parts and treatments. We will then report the procedures that characterized the experiment.

2.1 Experimental parts and treatments

The first part required participants to choose an effort level from a predefined set, which included six discrete levels in the closed interval $[0, 5]$. Effort, denoted by e, yielded costs $c(e)$ and output $Y(e)$. The top panel of Table 1 displays costs and outputs for each effort level.

Table 1: Effort levels

Effort level e	0	1	2	3	4	5
Cost $c(e)$	0	50	100	150	200	250
Output $Y(e)$	0	150	300	370	430	460
$Y(e) - c(e)$	0	100	200	220	**230**	210
$Y(e)/2 - c(e)$	0	25	**50**	35	15	-20

To operationalize effort costs (or disutility from work) as monetary expenditures is a well-established experimental method (see, for instance, Nalbantian and Schotter 1997). For the purpose of our study, a chosen-effort task has two important advantages compared to a real-effort task. First, it prevents personal variables—such as ability, motivation, and experience—to distort the effort. Second, it is gender neutral.

Before participants had to choose their effort, we manipulated two factors in a between-subjects design. The first factor is the payment scheme

[5] Such a payoff mechanism avoids wealth effects and hedging strategies that may emerge if subjects are paid for all decisions either sequentially or at the end of the experiment.

(*individual payment* versus *team payment*) and the second factor is the assignment method to the payment scheme (*random assignment* versus *self-selection*).

Concerning the first factor, participants were instructed that their payoff could be calculated according to two alternative payment schemes. In the *individual payment* scheme, participants earned the difference between their own output and their own effort costs, i.e., participant i's payoff was given by

$$Y(e_i) - c(e_i).$$

Conversely, the *team payment* scheme placed participants in 2-person teams, where partners equally shared the output and paid their effort costs individually. In addition, each team member was granted a fixed fee of 20 ECU.[6] Thus, participant i's payoff in the *team payment* scheme was given by

$$\frac{Y(e_i) + Y(e_j)}{2} - c(e_i) + 20.$$

The bottom panel of Table 1 shows that the payoff-maximizing effort in the *individual payment* scheme is $e = 4$. The same effort level also maximizes the *joint* payoff in teams (cooperative effort). Due to output sharing, the effort level that maximizes the individual payoff in teams is $e = 2$. It is worth noting that, as the production function is additively separable, in our team setting both the cooperative effort and the individual payoff-maximizing effort are independent of the partner's effort.

The second design factor manipulates whether participants were randomly assigned to one of the two payment schemes (*random assignment*) or could choose it (*self-selection*).

By crossing the two design factors, we produced four treatments: random assignment with individual payment (henceforth, RA-I treatment); random assignment with team payment (RA-T treatment); self-selection with individual payment (SS-I treatment); self-selection with team payment (SS-T treatment).

In the second part we asked all participants to subjectively assess the behavior of the subjects rewarded with *team payment* in the first part of the

[6]Without such an "entry fee", participants would have had no incentives to self-select into the team payment scheme. In practice, this fee can be interpreted as synergies in teams, which may compensate the incentive problems arising from output sharing.

experiment. Hence, participants in the random assignment treatments had to guess the behavior of subjects in the RA-T treatment, while participants in the self-selection treatments had to guess the behavior of subjects in the SS-T treatment. Beliefs were elicited by endowing participants with 100 tokens and asking them to allocate these tokens among the 6 possible effort levels. Participants were required to assign tokens to each alternative in a way that reflected the probability they attached to the event that the participants rewarded with *team payment* chose that alternative. We can think of each token as representing one percentage point.

We gave subjects proper incentives for accurate predictions using a probabilistic scoring rule that combines a quadratic scoring rule with a binary lottery procedure (Schlag and van der Weele, 2013; Hossain and Okui, 2013).[7] The rule was designed as follows. Define $r(e)$ as the percentage of participants rewarded with *team payment* that chose effort e and let $b_i(e)$ be the corresponding beliefs of participant i, with $\sum_{e=0}^{5} b_i(e) = 100$. Participant i's prediction error is calculated using the following loss function:

$$\ell_i = 0.005 \sum_{e=0}^{5} [b_i(e) - 100 \cdot r(e)]^2$$

where, the lower the loss function, the higher the accuracy of beliefs. Each participant received 200 ECU if her prediction error was smaller than a random number generated from a uniform distribution in $[0, 100]$ and earned 20 ECU otherwise. To dispel any doubts concerning the actual implementation of this procedure, in the instructions we used a verbal description of the rule and gave numerical examples. We also emphasized that it was in the participants' own interest to report their beliefs truthfully and that the more accurate their beliefs, the higher their chance of earning 200 ECU.

In the remaining three parts we collected data on additional personal characteristics. In Part 3 we measured participants' social value orientation using the six primary items of the Slider Measure by Murphy et al. (2011). For each item, participants faced nine options specifying different monetary allocations between themselves and another participant, and were asked to indicate their preferred one. Based on participant's choices, we computed the social value orientation angle (SVO). Participants were categorized into

[7] Schlag and van der Weele (2013) and Hossain and Okui (2013) independently developed an axiomatic characterization of this rule calling it "randomized quadratic scoring rule" and "binarized scoring rule", respectively. The main advantage of this rule is that it is incentive compatible irrespectively of the participants' risk preferences.

the following types: Altruist (SVO $>$ 57.15), Prosocial (22.45 $<$ SVO $<$ 57.15), Individualist (-12.04 $<$ SVO $<$ 22.45), Competitor (SVO $<$ -12.04). Instructions made clear that, if Part 3 was randomly selected for payment, each participant would be placed in a pair and randomly assigned to one of two roles (sender and receiver). Only one of the sender's items would then be paid out to both pair members, where each item had the same probability of being selected.

Part 4 estimates participants' trust using a binary trust game (McCabe and Smith, 2000), in which both the trustor and the trustee are endowed with 100 ECU. The trustor could either trust by passing all 100 ECU to the trustee, or pass nothing. In the former case, the 100 ECU were tripled to 300, so that the trustee had 400 ECU and the trustor had 0 ECU. Then, the trustee could either be trustworthy by sending 150 ECU back to the trustor, or send nothing and keep the entire 400 ECU. Being interested in measuring all participants' trust, the game was played using the strategy vector method. Participants were informed that, if Part 4 was randomly selected for payment, they would be randomly assigned to one of the two roles and paired with a participant in the opposite role. Their respective decision in the selected role would then be matched and the resulting payoff paid out.

Part 5 measures participants' risk attitudes using a multiple price list similar to that proposed by Holt and Laury (2002). Subjects faced ten pairwise comparisons. In each comparison they were asked to choose between a safe option that paid 150 ECU and a lottery that paid 210 ECU with probability p or 90 ECU with probability $(1-p)$. In each successive comparison, p increased by 10 percentage points, until finally the last decision involved no uncertainty. Subjects' risk preferences are inferred from the comparison at which they switch from the safe option to the lottery.[8] Instructions made clear that, if Part 5 was randomly selected for payment, only one of the ten decisions would be paid out, where each decision had the same probability of being selected.

Finally, participants were administered a post-experimental questionnaire asking them about (i) socio-demographic characteristics such as age and gender, (ii) their willingness to take risks in general, and (iii) their

[8]Consistent subjects should only switch once from the safe option to the lottery, and never back from the lottery to the safe option.

generalized trust attitude. The willingness to take risks was elicited using a non-incentivized question from the German Socio-Economic Panel, which required participants to self-assess their risk attitude using an 11-point Likert scale, ranging from 0 (unwilling to take risks) to 10 (fully prepared to take risks). Trust was measured using a well-known question from the World Value Survey, which read as follows: "Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?".

2.2 Procedures

The experiment was programmed in z-Tree (Fischbacher, 2007) and conducted in the experimental laboratory of the Frankfurt School of Finance & Management (Germany). The participants—students from different universities in Frankfurt—were recruited using the ORSEE software (Greiner, 2004). Upon entering the laboratory, they were randomly assigned to visually isolated computer terminals.

Each session consisted of the five parts explained in the previous section. The full sequence of events unfolded as follows. First, the instructions for Part 1 were distributed and read aloud in order to establish public knowledge. Participants received detailed information about the effort task and were familiarized with the payment schemes (*individual payment* or *team payment*).[9] Before making any choices, participants had to go through a series of control questions which tested their comprehension of the alternative payment schemes. Only after we ensured that everyone understood the instructions, subjects were either assigned to one of the two payment schemes (*random assignment*) or asked to choose their preferred one (*self-selection*).[10] Then the corresponding treatment (RA-I, RA-T, SS-I, or SS-T) started and subjects chose their effort levels. After all participants had made their decisions for Part 1, they received the instructions for Part 2. To mitigate potential demand effects, this procedure was maintained throughout

[9]To avoid exposing the design to the risk of undetectable experimenter demand effects, subjects in both assignment methods (random assignment and self-selection) were informed about the alternative payment schemes.

[10]In the SS-T treatment, teams were formed by mutual consent of being rewarded with team payment. If an odd number of subjects self-selected into the team payment, one of these subjects was randomly selected to be rewarded with individual payment. The selected participant was informed about the change in her payment scheme before she had to choose the effort level.

the experiment (i.e., the instructions for each new part were distributed and read after completion of the previous part).

Overall, we ran ten sessions with a total of 200 participants. At the end of each session, a randomly selected participant determined the part that was paid out by drawing one of five balls numbered 1 to 5 from an opaque bag. The outcome of the draw applied to all participants. Each session lasted less than 2 hours, including distribution and reading of the instructions as well as payment of money. Participants earned on average €24.10 (inclusive of a €5 show-up fee), ranging from a minimum of €7 to a maximum of €35.

3 Results

Table 2 displays summary statistics for effort levels, beliefs, and the personal characteristics collected in Parts 3, 4, and 5 of the experiment, and in the post-experimental questionnaire. For each assignment method, we report means and standard deviations for the subsample rewarded with *individual payment* (column labeled "Indiv."), for the subsample rewarded with *team payment* ("Team"), and for the pooled subsamples ("Full").

Table 2: Personal characteristics

	Random Assignment			Self-selection		
	Indiv.	Team	Full	Indiv.	Team	Full
Effort Level	3.86	3.14	3.50	3.92	3.42	3.67
Beliefs	3.32	3.19	3.25	3.06	3.44	3.25
Prosocial	0.62	0.70	0.66	0.66	0.70	0.68
Individualist	0.38	0.30	0.34	0.34	0.30	0.32
Trust	0.44	0.48	0.46	0.46	0.58	0.52
Trustworthiness	0.48	0.58	0.53	0.52	0.54	0.53
Self-assessed Trust	0.58	0.42	0.50	0.32	0.52	0.42
Switch in H&L	6.30	6.48	6.39	6.70	6.06	6.38
Self-assessed risk attitude	4.48	5.26	4.87	5.48	4.60	5.04
N	50	50	100	50	50	100

Note: Relative frequencies for prosocial, individualist, trust, trustworthiness and self-assessed trust; means for switch in H&L and self-assessed risk attitude.

Individual characteristics. We begin our analysis by verifying that the randomization of participants to assignment methods was effective. The choices in the social value orientation task reveal that a large fraction of subjects are classified as "prosocial" (66% and 68% in random assignment and self-selection, respectively), while the remaining participants are classified as "individualist". These proportions are in line with those reported in previous studies (see, among others, Murphy et al., 2011). About half of the participants trust by sending the entire endowment, and are trustworthy. Similar results are obtained when looking at the self-assessed trust attitude. According to a series of Fisher's exact tests, there are no significant differences in any of these individual characteristics between assignment methods (p-values equal to 0.881, 0.480, 1.000, and 0.321 for social value orientation, trust, trustworthiness, and self-assessed trust, respectively).

Risk attitudes are measured by the switching point in the Holt and Laury (H&L)'s multiple price list, which can range from 0 (if a participant is extremely risk lover and chooses the lottery in all pairwise comparisons) to 10 (if a participant is extremely risk averse and always chooses the safe option).[11] Hence, the higher the switching point, the higher the degree of risk aversion. On average, participants in both *random assignment* and *self-selection* are risk averse and switch around the sixth comparison. The same attitude toward risk emerges from the self-assessed measure.[12] According to two Wilcoxon rank sum tests, there are no significant differences in risk preferences between assignment methods (p-values equal to 0.619 and 0.658 for switching point and self-assessed risk attitude, respectively). The randomization assumption cannot therefore be rejected.

Determinants of self-selection. Table 3 shows the gender distribution across treatments and the p-values from proportion tests for the differences between relative frequencies. In *random assignment*, about 54% of females and 46% of males are rewarded with *team payment*. These frequencies are consistent with an equal split of females and males across payment schemes, which further guarantees that randomization worked well (p-value $\geq$ 0.414 for females and males). When participants can deliberately choose the payment scheme, females choose *team payment* more often than *individual pay-*

[11]Overall, thirteen participants switched more than once in the multiple price list. In such a case, we considered the first comparison in which subjects switched from the safe option to the lottery. Had we excluded inconsistent subjects or used the number of safe choices to measure risk attitudes, results would have remained unchanged.

[12]To make the two risk measures comparable, the self-assessed risk attitude was recoded such that higher numbers indicate higher degrees of risk aversion.

Table 3: Gender distribution across treatments

	Indiv. payment	Team payment	p-value for difference btwn Indiv. and team payment
(a) *Random assignment*			
Females	24 (46.15%)	28 (53.85%)	0.433
Males	26 (54.17%)	22 (45.83%)	0.414
(b) *Self-selection*			
Females	19 (38.00%)	31 (62.00%)	0.016
Males	27 (54.00%)	23 (46.00%)	0.424

Note: The table shows absolute (and relative) frequencies. Sample sizes are 52 females and 48 males in random assignment, and 50 females and 50 males in self-selection. p-values are from two-tailed proportion tests.

ment (62% vs. 38%, p-value = 0.016), while males evenly split across payment schemes (p-value = 0.423). Hence, we report the following result:

Result 1. *If given the possibility to choose the payment scheme, females are much more likely to choose the team payment scheme than the individual payment scheme, while males equally split across schemes.*

Next, we aim to shed light on the determinants of self-selection to team payment. Table 4 reports the results of several logit models in which the choice of the *team payment* is regressed on a series of relevant variables. More specifically, the table displays marginal effects, evaluated at the mean values of the independent variables.

Overall, our result show that beliefs—calculated by averaging all the possible effort levels, weighted for the corresponding expected probabilities—are a very important drive of self-selecting into team payment. As the own payoff is increasing in the partner's effort level, beliefs might be interpreted as the expected level of cooperation. It is thus not surprising to observe that a higher expected effort level raises the probability to choose *team payment.* The strong impact of beliefs holds true even when including a gender dummy in our specification. Interestingly, in line with the findings of Kuhn and Villeval (2015) and with the relative frequencies reported in Table 3, the results reported in the second column of Table 4 reveal that females self-select more often to *team payment* than males, *ceteris paribus.*

To verify whether beliefs equally affect the choice of both females and males or whether further personal characteristics affect differently the choice

Table 4: Determinants of choosing team payment

	All		Females	Males
Beliefs	0.279^{***}	0.297^{***}	0.328^{**}	0.271^{**}
	(0.090)	(0.093)	(0.130)	(0.126)
Prosocial	0.176	0.142	0.168	0.072
	(0.117)	(0.121)	(0.176)	(0.162)
Trust	0.056	0.062	-0.024	0.134
	(0.108)	(0.110)	(0.149)	(0.156)
Switch in H&L	-0.045	-0.047	-0.010	-0.084^{*}
	(0.031)	(0.031)	(0.043)	(0.046)
Female		0.197^{*}		
		(0.113)		
N	100	100	50	50

Note: The table displays marginal effects from logit regressions, evaluated at the mean values of the independent variables. Standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

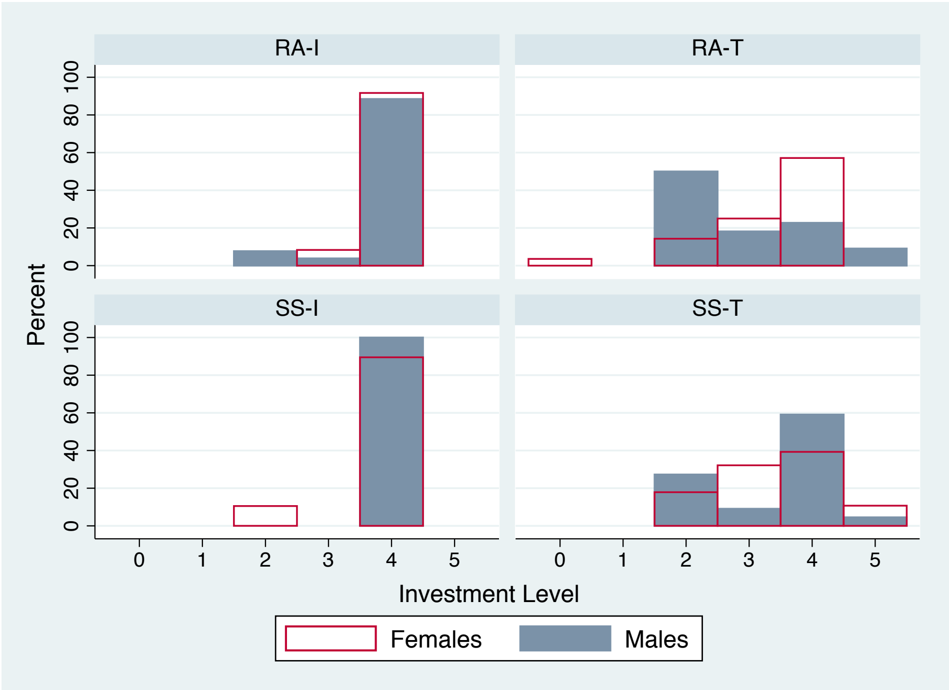
of men and women, we divide the SS-T sample by gender. The corresponding results are reported in the last two columns of Table 4. It turns out that the analysis is meaningful for men, whose choice to self-select into *team payment* is affected also by risk attitudes at the 5% level. In particular, the less risk averse males are, the more likely they are to choose team payment. This is quite intuitive as the payoff is deterministic with *individual payment*, but uncertain in teams. Social value orientation and trust, on the other hand, do not affect significantly either female or male choices.

Result 2. *Females are more likely than males to choose team payment.*

Result 3. *The choice of being rewarded with team payment is mainly driven by beliefs. In addition, male choices are affected by risk preferences.*

Determinants of effort levels. We now turn to the main purpose of the paper – that is if and how effort levels are affected by the payment scheme and the assignment method. Since teams are formed by mutual consent, if an odd number of participants self-selected into teams, one of them was randomly assigned to the alternative payment scheme. The change in payment scheme (from *team payment* to *individual payment*) was implemented for four participants (3 females and 1 males) and revealed to them before they could choose the effort level. As this change might have influenced the effort choice in an unpredictable way, we elect to follow a conservative approach

Figure 1: Distribution of Investment levels by treatment and gender



and exclude them from the analysis of effort levels. Hence, the sample size in SS-I reduces from 50 to 46 subjects.

Figure 1 and Table 5 show the distribution of effort levels across treatments, divided by gender. They show various things. First of all, irrespectively of the assignment method, the vast majority of participants choose the payoff-maximizing effort level of four (henceforth, first-best effort) when rewarded with *individual payment*, thus assuring that the payment scheme is clearly understood.

Table 5: Relative frequencies of effort levels by treatment and gender

	RA-I		RA-T		SS-I		SS-T	
e	Females (n=24)	Males (n=26)	Females (n=28)	Males (n=22)	Females (n=19)	Males (n=27)	Females (n=28)	Males (n=22)
0			3.57%					
1								
2		7.69%	14.29%	50.00%	10.53%		17.86%	27.27%
3	8.33%	3.85%	25.00%	18.18%			32.14%	9.09%
4	91.67%	88.46%	57.14%	22.73%	89.47%	100.00%	39.29%	59.09%
5				9.09%			10.71%	4.55%

Second, in RA-T the relative frequency of females choosing the cooperative (or first-best) effort level—that is, the effort that maximizes the team's joint payoff—is far higher than the corresponding frequency of males, and the difference is significant according to a Fisher exact test (57% vs. 23%, p-value = 0.021). Conversely, the relative frequency of females choosing the individually rational effort level—that is, the effort that maximizes the individual payoff, irrespectively of the effort level chosen by the partner—is far lower than the corresponding frequency of males (14% vs. 50%, p-value = 0.012).

Result 4. *In random assignment with team payment, females choose the cooperative effort level more often than males, who in turn are more likely to choose the individually rational effort level.*

Third, the comparison between RA-T and SS-T reveals that the frequency of females choosing the cooperative effort level drops from 57% to 39%, but the difference is insignificant. By contrast, there is a sharp positive effect of self-selection for males as the frequency increases from about 23% to 59% (Fisher exact tests; p-values equal to 0.285 and 0.031 for females and males, respectively).

Result 5. *Males (but not females) who prefer to be rewarded with team payment are more likely to choose the cooperative effort than those who are randomly assigned to the same payment scheme.*

To further examine treatment effects and investigate the impact of other personal characteristics on the choice of the first-best effort ($e = 4$) we run a series of logit models. Table 6 reports marginal effects, evaluated at the mean values of the independent variables, for all subjects (models (1) to (3)), and separately for females (models (4) and (5)) and males (models (6) and (7)). Intuitively, our findings suggest that subjects are more likely to choose the first-best effort when rewarded with individual payment than when rewarded with *team payment*.

Model (2) confirms the gender difference in the impact of self-selection, since the interaction term between SS-T and the gender dummy is significantly positive at the 5%-level. Adding the control variables in model (3) shows that the choice of the first-best level is mitigated by both the beliefs about the behavior of the partner and trust.

Table 6: Determinants of first-best effort

	Full			Females		Males	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
SS-I	0.544***	0.554***	0.521***	0.321**	0.403***		
	(0.104)	(0.102)	(0.079)	(0.134)	(0.127)		
SS-T	0.038	-0.068	-0.083	-0.125	-0.131	0.269***	0.098
	(0.064)	(0.080)	(0.071)	(0.090)	(0.089)	(0.093)	(0.101)
RA-I	0.401***	0.408***	0.355***	0.366***	0.382***	0.551***	0.434***
	(0.067)	(0.067)	(0.060)	(0.130)	(0.118)	(0.061)	(0.075)
Prosocial			-0.041		-0.022		-0.063
			(0.056)		(0.100)		(0.086)
Beliefs			0.178***		0.174***		0.195***
			(0.037)		(0.067)		(0.053)
Trust			0.144***		0.119		0.219***
			(0.048)		(0.077)		(0.075)
Switch			-0.011		-0.018		-0.024
			(0.013)		(0.022)		(0.020)
Male	-0.021	-0.120	-0.061				
	(0.057)	(0.073)	(0.070)				
Male x SS-T		0.244**	0.134				
		(0.111)	(0.104)				
N	196	196	196	99	99	70	70
Pseudo R2	0.2392	0.2576	0.3989	0.1881	0.2754	0.2425	0.4946

Note: The table displays marginal effects from logit regressions, evaluated at the mean values of the independent variables. The baseline treatment is RA-T. In the last two colummns, the SS-I treatment dummy is omitted and 27 males are excluded from the subsample as all of them chose the first-best effort level. Standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Result 6. *Subjects are more likely to choose the first-best effort level when they believe that the partner behaves more cooperatively.*

In addition, model (3) shows that the interaction term between the SS-T and gender becomes insignificant when adding the personal control variables. The reason can be seen by comparing the female and male subsamples in models (5) and (7), respectively. For both genders, beliefs are highly positive and significant. Trust, however, is positive at the 1%-level for males, but insignificant for females. Thus, part of the positive impact of self-selecting into teams for males is absorbed by beliefs and trust in model (3), because the levels for trust and beliefs are higher for males that self-select into teams compared to random assignment. Finally, in line with the non-parametric test reported above, the SS-T-dummy is significantly positive for males and insignificant for females.

4 Conclusions

Preceding laboratory experiments have shown that the degree of moral hazard in teams depends on multiple factors such as team composition, personality and the behavior of other participants. We contribute to this literature by distinguishing between two treatments. In the first treatment, subjects are randomly assigned to individual compensation or revenue-sharing in teams. In the second treatment, they can self-select in one of the two payment schemes. To isolate the impact of self-selection, we consider a chosen effort-setting instead of real effort, and we neglect synergies in teams by assuming that the production function is additively separable. Focusing on gender differences, our main results are threefold: First, with random assignment, females choose the cooperative effort level more often than males. Second, females are more likely to self-select to teams. Third and most interesting in our view, self-selection increases the percentage of males choosing the cooperative effort from 23% to 59%, while there is no positive impact of self-selection for females. For both genders, choosing teams and effort levels are largely driven by expectations on the partner's behavior.

Our finding that self-selection matters a lot for males but not so for females seems important for the optimal process on team compositions in practice. Considering chosen effort with an additively separable production function allows to identify the pure impact of self-selection on moral hazard, but comes at the cost of neglecting e.g. team heterogeneity, synergies and gender composition in teams.

Next, our regression analyses on effort levels show that higher efforts are not only driven by self-selection, gender and beliefs on the partner's effort, but also by the level of trust. The impact of personality suggests that, in addition to random assignment and self-selection, it is interesting to consider a third allocation mechanism, where the experimenter assigns subjects to either individual or team payment. The idea is to start with simple games such as the trust game, and to make the assignment decisions dependent on the behavior in these games. Given findings that the behavior in simple experiments can be used as a predictor for outcomes long-term real life-situations, it would be interesting to see whether such a mechanism increases the overall output.

References

Alchian, A. A. and Demsetz, H. (1972). Production, information costs, and economic organization. *American Economic Review*, 62(5):777–795.

Apesteguia, J., Azmat, G., and Iriberri, N. (2012). The impact of gender composition on team performance and decision making: Evidence from the field. *Management Science*, 58(1):78–93.

Bäker, A. and Mertins, V. (2013). Risk-sorting and preference for team piece rates. *Journal of Economic Psychology*, 34:285–300.

Bäker, A. and Pull, K. (2016). Who is attracted by teamwork? Evidence of multidimensional sorting from a real-effort experiment. *German Journal of Human Resource Management*, forthcoming.

Ball, S. and Eckel, C. C. (1998). The economic value of status. *The Journal of Socio-Economics*, 27(4):495–514.

Bradley, B. H., Klotz, A. C., Postlethwaite, B. E., and Brown, K. G. (2013). Ready to rumble: how team personality composition and task conflict interact to improve performance. *Journal of Applied Psychology*, 98(2):385–392.

Cadsby, C. B., Song, F., and Tapon, F. (2007). Sorting and incentive effects of pay for performance: An experimental investigation. *The Academy of Management Journal*, 50(2):387–405.

Chao, H. and Croson, R. T. (2013). An experimental comparison of incentive contracts in partnerships. *Journal of Economic Psychology*, 34:78–87.

Corgnet, B., Hernán-González, R., and Schniter, E. (2015). Why real leisure really matters: incentive effects on real effort in the laboratory. *Experimental Economics*, 18(2):284–301.

Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2):448–474.

Delfgaauw, J., Dur, R., Sol, J., and Verbeke, W. (2013). Tournament incentives in the field: Gender differences in the workplace. *Journal of Labor Economics*, 31(2):305–326.

Dohmen, T. and Falk, A. (2011). Performance pay and multidimensional sorting: Productivity, preferences, and gender. *American Economic Review*, 101(2):556–590.

Eckel, C. C. and Grossman, P. J. (2008). Men, women and risk aversion: experimental evidence. In Plott, C. and Smith, V., editors, *Handbook of Experimental Economics Results, vol. 1*, chapter 113, pages 1061–1073. New York: Elsevier.

Encinosa III, W. E., Gaynor, M., and Rebitzer, J. B. (2007). The sociology of groups and the economics of incentives: Theory and evidence on compensation systems. *Journal of Economic Behavior & Organization*, 62(2):187–214.

Eriksson, T. and Villeval, M. C. (2008). Performance-pay, sorting and social motivation. *Journal of Economic Behavior & Organization*, 68(2):412–421.

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

Fischbacher, U. and Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review*, 100(1):541–556.

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

Gächter, S. (2007). Conditional cooperation: Behavioral regularities from the lab and the field and their policy implications. In Frey, B. S. and Stutzer, A., editors, *Economics and Psychology. A Promising New Cross-Disciplinary Field*, pages 19–50. The MIT Press 2007.

Gächter, S. and Thöni, C. (2005). Social learning and voluntary cooperation among like-minded people. *Journal of the European Economic Association*, 3(2-3):303–314.

Glew, D. J. (2009). Personal values and performance in teams: An individual and tam-level analysis. *SAGE Journals*, 40(6).

Greiner, B. (2004). An online recruitment system for economic experiments. In Kremer, K. and Macho, V., editors, *Forschung und wissenschaftliches Rechnen 2003*, pages 79–93. Göttingen: Gesellschaft für Wissenschaftliche Datenverarbeitung.

Hamilton, B. H., Nickerson, J. A., and Owan, H. (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.

Holmström, B. (1982). Moral hazard in teams. *The Bell Journal of Economics*, 13(2):324–340.

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

Hossain, T. and Okui, R. (2013). The binarized scoring rule. *The Review of Economic Studies*, 80(3 (284)):984–1001.

Ivanova-Stenzel, R. and Kübler, D. (2011). Gender differences in team work and team competition. *Journal of Economic Psychology*, 32(5):797–808.

Keser, C. and Montmarquette, C. (2011). Voluntary versus enforced team effort. *Games*, 2(3):277–301.

Keser, C. and van Winden, F. (2000). Conditional cooperation and voluntary contributions to public goods. *The Scandinavian Journal of Economics*, 102(1):23–39.

Kosfeld, M. and von Siemens, F. A. (2009). Worker self-selection and the profits from cooperation. *Journal of the European Economic Association*, 7(2-3):573–582.

Kuhn, P. and Villeval, M. C. (2015). Are women more attracted to cooperation than men? *The Economic Journal*, 125(582):115–140.

Mathieu, J. E., Tannenbaum, S. I., Donbach, J. S., and Allgier, G. M. (2014). A review and integration of team composition models. *Journal of Management*, 40(1):130–160.

McCabe, K. A. and Smith, V. L. (2000). A comparison of nave and sophisticated subject behavior with game theoretic predictions. *Proceedings of the National Academy of Sciences*, 97(7):3777–3781.

Mohnen, A., Pokorny, K., and Sliwka, D. (2008). Transparency, inequity aversion, and the dynamics of peer pressure in teams: Theory and evidence. *Journal of Labor Economics*, 26(4):693–720.

Murphy, R. O., Ackerman, K. A., and Handgraaf, M. J. J. (2011). Measuring social value orientation. *Judgment and Decision Making*, 6(8):771–781.

Nalbantian, H. R. and Schotter, A. (1997). Productivity under group incentives: An experimental study. *American Economic Review*, 87(3):314–341.

Newhouse, J. P. (1973). The economics of group practice. *The Journal of Human Resources*, 8(1):37–56.

Nosenzo, D. and Tufano, F. (2015). Entry or exit? the effect of voluntary participation on cooperation. Discussion Papers 2015-20, The Centre for Decision Research and Experimental Economics, School of Economics, University of Nottingham.

Pizzini, M. (2010). Groupbased compensation in professional service firms: An empirical analysis of medical group practices. *The Accounting Review*, 85(1):343–380.

Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature*, 37(1):7–63.

Schlag, K. H. and van der Weele, J. J. (2013). Eliciting probabilities, means, medians, variances and covariances without assuming risk neutrality. *Theoretical Economics Letters*, 3(1):38–42.

Vyrastekova, J., Onderstal, S., and Koning, P. (2012). Self-selection and the power of incentive schemes: an experimental study. *Applied Economics*, 44(32):4211–4219.