DISCUSSION PAPER SERIES

IZA DP No. 2001

Performance Pay and Multi-dimensional Sorting: Productivity, Preferences and Gender

Thomas Dohmen
Armin Falk

March 2006

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Performance Pay and Multi-dimensional Sorting: Productivity, Preferences and Gender

Thomas Dohmen
IZA Bonn

Armin Falk
IZA Bonn and University of Bonn

Discussion Paper No. 2001
March 2006

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
Email: iza@iza.org

IZA Discussion Paper No. 2001
March 2006

ABSTRACT

Performance Pay and Multi-dimensional Sorting: Productivity, Preferences and Gender

This paper studies the impact of incentives on worker self-selection in a controlled laboratory experiment. In a first step we elicit subjects' productivity levels. Subjects then face the choice between a fixed or a variable payment scheme. Depending on the treatment, the variable payment is either a piece rate, a tournament or a revenue-sharing scheme. We elicit additional individual characteristics such as subjects' risk attitudes, measures of self-assessment and overconfidence, social preferences, gender and personality. We also elicit self-reported measures of work effort, stress and exhaustion. Our main findings are as follows. First, output is much higher in the variable pay schemes (piece rate, tournament, and revenue sharing) compared to the fixed payment scheme. Second, this difference is largely driven by productivity sorting. On average, the more productive a worker is, the more likely he self-selects into the variable pay scheme. Third, relative self-assessment and overconfidence affect worker self-selection, in particular into tournaments. Fourth, risk averse workers prefer fixed payments and are less likely to sort into variable pay schemes. Fifth, people endowed with social preferences are less likely to sort into tournaments. Sixth, variable pay schemes attract men more than women, a difference that is partly explained by gender-specific risk attitudes. Seventh, self-selection is also affected by personality differences. Finally, reported effort is significantly higher in all variable pay conditions than in the fixed wage condition. In sum, our findings underline the importance of multi-dimensional sorting, i.e., the tendency for different incentive schemes to systematically attract people with different abilities, preferences, self-assessments, gender and personalities.

JEL Classification: M52, M55, J00, J3, J33, J31, J16, J22, J24 C91, D81

Keywords: personnel economics, sorting, incentives, productivity, ability, piece rates, tournament, revenue sharing, risk preferences, overconfidence, gender, experiment

Corresponding author:

Armin Falk
IZA
P.O. Box 7240
D-53072 Bonn
Germany
Email: falk@iza.org

1 Introduction

Typically the rationale for providing incentive schemes is to align the interests of principals and agents in the presence of a contract enforcement problem. This view neglects the importance of worker self-selection, i.e., the possibility that agents with different individual characteristics feel attracted by different pay schemes and therefore self-select into firms and organizations. In the presence of self-selection the overall output effects of different incentive systems are likely to depend not only on the incentive effect *per se* but also on the endogenous composition of the workforce. Mostly due to data limitation and confounding factors, relatively little is known empirically about the nature of this selection process. This paper therefore explores the driving forces of self-selection in a controlled laboratory environment. In particular we address the following questions: Which personal characteristics provoke workers to self-select into variable instead of fixed pay contracts? Is the sorting decision driven by individual productivity differences and if so, do other relevant characteristics like risk aversion, overconfidence, social preferences, gender or personality shape the selection process? How does the composition of the workforce differ when firms offer either fixed wages or variable payments in the form of piece rates, tournaments or revenue sharing?

The idea of the experiment is to first elicit subjects' individual productivity levels. Subjects then face the choice between a variable and a fixed payment scheme. We observe which payment mode they prefer and how much they work. Finally, we elicit further individual characteristics that may be relevant for the sorting decision. Among them are subjects' risk attitudes, self-assessment and overconfidence, social preferences, gender and personality. In addition we elicit self-reported measures of work effort, stress and exhaustion. The work task consists of multiplying one and two digit numbers and is characterized by a substantial degree of heterogeneity in productivity. We study three treatment conditions, which are characterized by different variable pay schemes. This allows us to study the sorting patterns when the choice is between a fixed payment and either a piece rate, a tournament, or a revenue-sharing scheme. These three forms of variable pay constitute the most important prototypical forms of explicit performance incentives. Since the treatments are exactly identical except for the alternative variable pay scheme, our design allows us to study different sorting patterns as a response to these different pay schemes in a uniform and comprehensive framework.

Our main results reveal the importance of multi-dimensional sorting. First of all there is substantial productivity sorting in all three treatments. When facing the alternative between variable and fixed payments, more productive workers systematically prefer the variable pay. This holds regardless of whether the latter is offered as a piece rate, a tournament or a revenue-sharing scheme. Second, relative self-assessment and overconfidence affect sorting into tournaments, which is sensible as payments in tournaments depend on relative performance. The better subjects think they perform relative to others and the more they overestimate their relative performance, the more likely they are to enter tournament competition. Third, controlling for individual productivity, risk attitudes play an important role in the sorting decision: relatively risk averse workers prefer receiving fixed wages, while subjects who sort into variable pay turn out to have a relatively high risk tolerance. This finding reflects the fact that the fixed payment yields a safe payoff whereas earnings in the variable pay are uncertain and therefore risky. Our fourth result concerns social preferences: we find that tournaments seem to attract relatively selfish individuals. This may be driven by the fact that providing effort in tournaments imposes a negative externality on the competitors and that final payoffs in the tournament are quite unequal. These features are not present under piece rates or revenue sharing, and we observe no sorting based on social preferences in these schemes. In our fifth result we show that women are less likely to select into variable pay schemes than men. This is mostly explained by differences in risk attitudes between men and women. We also show that personality is relevant for self-selection. Interestingly, different personality traits are relevant for men and women. On top of the observed sorting patterns we see that self-reported effort varies significantly with different work incentives. In comparison to those working under fixed wages, subjects working under variable pay schemes report significantly higher effort levels as well as higher levels of stress and exhaustion.

The literature on optimal incentives has shown that characteristics of the production processes and the information structure affect the structure of optimal employment contracts.[1] Our results indicate that organizations should, in addition, take the interaction of incentives and multi-dimensional sorting seriously when deciding on the design of the incentive system. This is particularly important because individual characteristics that are decisive for the selection process and for the performance of the organization

[1] Early work (e.g., Stiglitz, 1975) focused on the role of monitoring costs and imperfect information about individuals' abilities. Implications for the choice between piece-rate contracts and time-rate contracts in the presence of monitoring costs have also been amply studied (see, e.g., Brown, 1990, 1992, 1994; Goldin, 1986; Parent, 1999; Pencavel, 1977). Lazear and Rosen (1981) have proposed rank order tournament as optimal incentive contracts when reliable monitors of effort are too costly. Optimal multiperiod incentive schemes have been considered in another strand of the literature, e.g., Laffont and Tirole, 1988, which also highlights the role of future commitment (see Baron and Besanko, 1984; Gibbons, 1987; Kanemoto and MacLeod, 1992). For evidence on the interplay between job characteristics and the incidence of particular compensation contracts see also MacLeod and Parent (1999).

are typically unobservable to firms during the hiring process. To illustrate, consider fund managers working under highly variable and competitive pay schemes. Given the observed sorting patterns we would expect that relatively risk tolerant and overconfident managers are attracted. This is likely to have consequences for the portfolio strategy and thus the performance of the fund. Likewise, introducing variable pay in certain jobs that are characterized by fixed wage schemes, such as the public sector, will change the composition of the workforce and consequently the output produced by this workforce. To the extent that firms, even when operating in similar environments, have different preferences regarding the composition of their workforce, our results offer an explanation for why firms install different remuneration schemes. Our results also suggest that firms may use incentive schemes as screening devices to attract particular types of workers (Salop and Salop, 1976).

Showing the relevance of sorting also underlines the methodological difficulties associated with testing contract theory with field data (Prendergast, 1999; Chiappori and Salanié, 2003). Comparing output under performance pay schemes to output when remuneration is independent of effort, it is often hard to determine whether higher output under the former is due to incentives or sorting. This point has been forcefully made in the theoretical analysis by Lazear (1986) and by Lazear (2000) in a case study of a firm that changes from fixed wages to piece rates. This case study shows that changing the incentive system gives rise to a substantial change in the composition of the workforce. Ignoring this selection effect would imply a dramatic overestimation of the incentive effect of piece rates. Our results confirm this conclusion about productivity sorting for piece rates, but also for tournaments and revenue sharing. Moreover, our results point to another potential confounding factor in testing contract theory: preference and self-assessment sorting. It is well known that optimal contracts depend on risk preferences. More recently it has also been argued that optimal contracting depends on social preferences, i.e., on the composition and interaction of selfish and reciprocal agents (see, e.g., Fehr, Klein and Schmidt, 2005; Englmaier and Wambach, 2005; Grund and Sliwka, 2005). In light of our findings, researchers interested in testing these theories should be aware of the fact that the composition of both risk and social preferences in a given pool of agents is likely to be endogenous.

Ruling out this kind of endogeneity is exactly one rationale for conducting laboratory experiments. In the lab it is easy to implement random treatment assignment in order to rule out sorting and to get unbiased estimates of the incentive effects of different incentive schemes. In this way, experiments have produced valuable and indispensable knowledge about the incentive effects of different incentive schemes.[2] However, by focusing on the incentive effect the literature has lost sight of the sorting issue, which is equally important for an understanding of the effectiveness of incentive pay. The present experiment shows that experiments can be used not only to rule out selection effects with random assignment but also to study sorting in a controlled way. Only recently have experimenters begun to run experiments in a similar vein. Sorting has been studied, e.g., in a market entry game (Camerer and Lovallo, 1999), in simple bargaining games (Oberholzer-Gee and Eichenberger, 2004; and Lazear, Malmendier and Weber, 2005), the gift-exchange game (Eriksson and Villeval, 2004) or the prisoner's dilemma game (Bohnet and Kübler, 2004). More related to our paper is Cadsby et al. (2005), who study sorting outcomes when the alternative is between piece rates and fixed wages, and Eriksson, Teyssier and Villeval (2005) who show that effort variability in tournaments is lower when agents can decide whether to work under piece rates or under tournament incentives. This is also the choice that subjects in the study by Niederle and Vesterlund (2005) face. Based on the finding that women perform worse in the presence of men in competitive environments (Gneezy and Rustichini, 2004) they study whether women shy away from competition. They find that women are less willing to compete in tournaments compared to men when the alternative is to work under piece rates. As mentioned above, this is similar to our finding that women are less likely to select into variable pay than men when the alternative is a fixed payment. In this sense sorting offers a possible channel for gender differences in occupational choice, career choice and ultimately for the existence of the gender wage gap.

The paper is organized as follows. The next section describes the experiment in some detail. Section 3 presents the results. We first discuss the output effects of different incentive schemes. Then we present evidence on the importance of sorting with respect to productivity, relative self-assessment, overconfidence, risk preferences, social preferences, gender and personality. Finally, we discuss the effect of incentives on the provision of effort. Section 4 concludes.

[2] Using random treatment assignment, tournament incentives have been studied, e.g., by Bull, Schotter and Weigelt (1987), Schotter and Weigelt (1992), Falk and Fehr (2002) and Harbring and Irlenbusch (2003). The lab evidence on tournaments is complemented by field studies on corporate tournaments (Bognanno, 2001), tournaments in agricultural production (Knoeber and Thurman, 1994) and sports tournaments (e.g., Ehrenberg and Bognanno, 1990; Fernie and Metcalf, 1999; and Sunde, 2003). The incentive effects of piece rates have been experimentally investigated, e.g., by Bull, Schotter and Weigelt (1987) and van Dijk, Sonnemans and van Winden (2001), while team incentives have been studied, e.g., by Nalbantian and Schotter (1997). The impact of incentives has also been studied in field experiments, e.g., Bandiera et al. (2005) and Nagin et al. (2002).

2 An Experimental Approach to the Study of Incentives and Multi-Dimensional Sorting

For studying how individual characteristics affect the sorting decision into different incentive schemes, the ideal data set combines knowledge of individual productivity and personal characteristics along with direct observation of the selection decision in a well defined environment. Such data are difficult to obtain in the field. First, it is often unclear, which kind of incentives actually prevail in an organization. Suppose, for example, that the researcher observes that a firm has established piece-rate contracts. This does not exclude the possibility that workers are in addition confronted with team incentives like profit sharing or compete against each other in promotion tournaments. Moreover, implicit contracts and repeated game effects may create work incentives even in the absence of explicit performance incentives (MacLeod and Malcomson, 1989, 1998). Thus, workers may be exposed to a mix of explicit and implicit incentives, which hinders a clear characterization of the work environment. Second, individual output measures are often not available or are fraught with measurement error. Third, individuals' characteristics and preferences are typically not observed. This holds for productivity measures but even more so for personal attributes like risk aversion, social preferences or overconfidence. Fourth, it is only appropriate to interpret policy changes in firms as natural experiments if these changes are exogenous, which is always debatable. Finally, policy changes need time to affect the endogenous composition of the workforce and it is not obvious what time frame the researcher should consider. Allowing too little time for sorting to take place, will lead to an underestimation of the sorting effect. Waiting too long, however, increases the likelihood that other factors besides the change in the incentive scheme will affect the sorting process.

We think that experiments offer a valuable tool for studying incentives and sorting in a controlled environment, complementing the evidence generated by observational field studies in an informative way. In the lab, it is possible to precisely define the material incentives upon which subjects can base their sorting decision. It is further possible to elicit measures of individual productivity with little measurement error as well as individual characteristics and preferences. Furthermore we can guarantee the absence of a mix of different implicit or explicit incentives. Finally, since the sorting decision takes place immediately, timing is not an issue.

2.1 The Work Task

The work task implemented in our experiment consists of multiplying one- and two-digit numbers. This "real effort" task implies that subjects have to actually work[3] and are to some extent uncertain about their productivity and the productivity of others. This is a realistic feature of most work tasks and leaves room for sorting according to (relative) self-assessment. As a task, multiplying numbers is also well suited for our purposes because it requires no previous knowledge, is easy to explain, and guarantees a sufficient degree of heterogeneity in productivity. Moreover, this task is a relatively good proxy for general cognitive ability, and in light of recent neuroscience evidence, learning effects during the experiment are expected to be small (Roth, 2001). Depending on the chosen numbers, the difficulty level of multiplying one- and two-digit numbers varies quite a bit. This has to do with the fact that different problems require different usages of working memory. In particular, we distinguish between five different degrees of difficulty.[4] As we will see below, solving more difficult problems is more time-consuming.

All problems were presented to subjects on computer screens. They could type their answer in a box and confirm it by clicking an "OK"-button with their mouse. Having entered the answer, a subject was informed whether or not the solution was correct. If it was correct, a new problem appeared instantaneously on the screen (except in steps 1 and 2 of the experiment where only one problem had to be solved, see below). If the answer was wrong, subjects had to tackle the same problem again until the correct solution was entered. We forced subjects to solve a problem before a new question appeared on the screen in order to prevent subjects from guessing and searching for "easy" problems. A subject was always informed about the cumulative number of problems he had answered correctly.

2.2 Design and Treatments of the Experiment

In order to study how individual characteristics affect the sorting decision into different incentive schemes, we implemented an experiment that includes 12 steps (see Figure A.1 in the Appendix). Subjects were informed at the beginning that they would go through

[3] This is in contrast to most economic labor market experiments that mimic effort costs by requiring subjects to choose a number, with higher numbers costing more money. Other real effort experiments include, e.g., Fahr and Irlenbusch (2000) who have subjects crack walnuts, van Dijk, Sonnemans and van Winden (2001) who asked subjects to perform cognitively demanding tasks on the computer, Gneezy et al. (2003) who had subjects solve mazes at the computer and Falk and Ichino (2006) who asked subjects to stuff letters into envelopes.

[4] Examples for the five levels of difficulty are: Level 1: $11 \cdot 9 =$???; Level 2: $3 \cdot 32 =$???; Level 3: $6 \cdot 43 =$???; Level 4: $4 \cdot 68 =$???; Level 5: $7 \cdot 89 =$???.

different steps, but they did not know what these steps would involve. The first three steps are designed to elicit three different measures of individual productivity. In the first step, all subjects were asked to calculate one multiplication problem as fast as possible. The problem that they were confronted with on the computer screen had a degree of difficulty of 4. No payment was involved. The time that elapsed before the correct answer was entered is our first productivity indicator (Productivity Indicator 1).

The second productivity measure is basically the same as the first, except that this time subjects were paid for being fast. Again they were asked to calculate one problem with degree of difficulty 4 as fast as possible. This time, they were endowed with 30 seconds of time to solve the question, each second being worth 5 points. Subjects were told that 5 points would be subtracted from this endowment for each second they needed for solving the problem. This means, e.g., that a subject who answered the question after 15 seconds earned 75 points while someone who needed 22 seconds received only 40 points, etc. Earnings for subjects who did not come up with the correct answer in 30 seconds were zero.

Our third measure of an individual's productivity (Productivity Indicator 3) is the number of problems that a subject solved when working for five minutes for a piece rate of 10 points per correct answer. Each subject went through the exact same sequence of problems. We implemented a stratified sampling design of questions, i.e., each block of 10 problems had the following structure in terms of difficulty: one problem of degree 1, one problem of degree 2, two problems of degree 3, four problems of degree 4, and two problems of degree 5. The sequence of questions within a block of ten questions was random.

The three productivity indicators measure different aspects of individual productivity. Productivity Indicator 1 measures willingness and ability to answer a problem quickly. Since no stakes are involved, it also captures an element of intrinsic motivation. Productivity Indicator 2 measures how fast subjects answer questions when they are under some pressure, which resulted because they were paid for being fast and because they saw the time that remained for answering the question on the screen. Productivity Indicator 3 measures output under a different incentive scheme and allows much more time. It is therefore a good proxy for subjects' productivity and perseverance. The latter is relevant since the work task after the sorting decision involves 10 minutes of problem solving. In section 3 we will therefore predominantly use Productivity Indicator 3.

In step 4 we asked subjects to subjectively assess how hard they had worked in the five minute working time in step 3. In particular, we asked the following three questions: How much effort did exert? How stressed did you feel? How exhausted did you get? Answers to these questions were given on a seven point scale, where the value 1 means 'not at all' and the value 7 means 'very much'. Then, in step 5, we asked subjects to assess their performance in step 3 relative to the performance of the other 19 participants in their session. We are interested in this assessment to find out whether it affects the sorting decision (in particular into tournaments) and whether selection into variable pay schemes is associated with overconfidence. The question subjects had to answer reads as follows: How many of the other 19 participants solved more question than you did? Subjects had an incentive to answer the question as accurate as possible. For a correct estimate they received 100 points, for a deviation of plus or minus one from the correct number they received 50 points, and otherwise they received zero points. Subjects were informed about their true rank in the distribution not until the very end of the experiment.

Step 6 is the actual sorting decision. Subjects were informed that they were to work for ten minutes on the same work task as before, i.e., multiplying one-digit and two-digit numbers with a similar degree of difficulty. Before they started to work, they were offered the choice between a variable pay contract and a fixed-payment contract. The chosen contract determined how they were paid for the output they produced later in the 10-minute work period. In each of our three treatments, the fixed-payment contract, w_F, defined by

$$w_F = \alpha, \quad \alpha > 0, \tag{1}$$

guaranteed the payment of $\alpha = 400$ points independent of output x, the number of correctly answered problems. It was made clear to subjects that they would receive 400 points independent of whether they solved a few, many, or no problems at all. The only requirement for receiving the 400 points was that they had to stay in the lab.

The type of variable pay scheme offered as an alternative to the fixed wage defines each of our three treatments. We study piece-rate, tournament or revenue-sharing contracts. These three forms of variable pay constitute the most important prototypical forms of explicit performance incentives. In the piece-rate treatment, the alternative contract paid a piece rate, β, of ten points per correct answer, just as in step 3. Remuneration of subject i according to the piece-rate contract, w_{PR}, is given by

$$w_{PR} = \beta \cdot x_i, \quad \beta > 0. \tag{2}$$

In the tournament treatment, a subject i could choose to compete in a two-person tournament, in which the opponent j was randomly chosen among all subjects who had also opted for the tournament. Among the two competitors, the subject who had solved more problems at the end of the 10-minute work period won the tournament and received the winner prize, γ, of 1,300 points. The loser received zero points. If both competitors had solved the same number of problems, the winner was determined by a random draw. The tournament contract is given by

$$w_T = \begin{cases} \gamma, & \text{if } x_i > x_j, \quad i \neq j, \quad 0 < \alpha < \gamma \\ 0.5 \cdot \gamma + 0.5 \cdot 0, & \text{if } x_i = x_j, \\ 0, & \textit{otherwise} \end{cases} \tag{3}$$

Subject i was informed about opponent j's output only after the working time of ten minutes was over. If an odd number of subjects had selected into the tournament, one randomly chosen subject's output was used a second time to determine the score of the unmatched subject's opponent. If only one subject opted for the tournament (which did not happen), no tournament was implemented and the subject was informed that he or she would be compensated according to the fixed-payment contract w_F. Subjects were informed about these details prior to their sorting decision.

In the revenue-sharing treatment, subjects could choose to work for a sharing contract as an alternative to the fixed wage contract w_F . Two subjects who opted for this compensation were randomly matched and formed a team. The team received a piece rate of $\beta = 10$ points for each correctly answered problem. A team's revenue was then divided equally among the two team members. The compensation in the revenue-sharing treatment w_{RS} is hence given by

$$w_{RS} = \beta \cdot \frac{x_i + x_j}{2}. \tag{4}$$

Again, the output of the other team member j was disclosed only after the end of the 10 minute working time. If only one subject or an odd number of subjects decided to work under revenue-sharing incentives, the same rules as under the tournament treatment applied.

Right after the sorting decision but before the actual working time began, we asked all subjects in step 7 how they would have decided if the fixed payment had been different. In particular, subjects had to indicate whether they would prefer the treatment-specific variable pay or the fixed payments of {50, 100, 150, ..., 800} points. These hypothetical choices reveal valuable information about sorting patterns at more and less attractive fixed payment alternatives.

Step 8 is the 10 minutes working time, during which subjects worked under their preferred contractual terms, i.e., either for a fixed payment of 400 points or for the respective variable pay. At the end of the working time, we notified subjects about their earnings, and we disclosed the competitor's output to tournament participants and the partner's output to team members. In step 9 we asked subjects to inform us on a seven point scale about effort, stress and exhaustion in exactly the same way as in step 4.

In the remaining three steps, we collected data on additional personal characteristics. In step 10 we elicited subjects' social preferences with the help of a simple trust game (similar to Berg et al., 1995). Each subject played a 2-player, sequential trust game. Both players received an endowment of 120 points. The first mover could transfer any amount {0, 20, 40, 60, 80, 100, 120} to the second mover. Any transfer was tripled. The second mover could then send back any amount between zero and 480. To elicit information about player types we used the contingent response method, i.e., second movers had to indicate for each of the seven possible transfer levels how much they wanted to transfer back to the first mover, before they knew the actual transfer. This is an incentive compatible way to elicit social preferences since any decision is potentially payoff-relevant. In order to be able to classify each subject, everybody had to play both roles, first and second mover. After all choices had been made, pairs of subjects were formed by random matching and the roles of first and second movers within a pair were assigned by a random draw. The players' choices were then implemented, and subjects were paid accordingly.

Step 11 elicits subjects' risk preferences using simple lottery choices, similar to Holt and Laury (2002). Participants in our experiment were shown a table with 15 rows. In each row they had to decide whether they preferred a safe option or playing a lottery. In the lottery they could win either 400 points or 0 points with 50 percent probability. The lottery was exactly the same in each row, but the safe option increased from row to row. In the first row, the safe option was 25 points; in the second it was 50 points, and so on up to 375 points in row 15. After a subject had made a decision for each row, it was randomly determined which row became relevant for payment. This procedure guarantees that each decision was incentive compatible. If subjects have monotonous preferences, they prefer the lottery up to a certain level of the safe option, and then switch to preferring the safe option in all subsequent rows of the choice table. The switching point informs us about a subject's risk attitude.

In the final step 12, we elicited subjects' risk attitudes in an alternative way, namely by asking individuals to indicate their willingness to take risks in general on an eleven-point scale, with zero indicating complete unwillingness to take risks, and ten indicating complete willingness to take risks. We use the same wording of the question as in the 2004 wave of the German Socioeconomic Panel (GSOEP), a representative panel survey of the resident population of Germany.[5] Dohmen et al. (2005) have validated the behavioral relevance of this general risk question in a field experiment with a representative subject pool of 450 individuals. They conclude that the survey risk measure is a good predictor of risky choices with real money at stake. Using data on roughly 22,000 individuals, they also show that answers to the general risk question predict behavior in many life contexts better than lottery choices do.

We also gathered questionnaire data on socioeconomic characteristics (including gender, age, nationality, marital status, and parents' education), on educational achievement (grades and major fields of study on university-entrance examination (Abitur), high-school graduation year, and last mathematics grade in high-school). Subjects also completed a verbal IQ-test, the so called MWT-A (Lehrl et al., 1991), and a personal attitudes test developed by Hermann Brandstätter (see Brandstätter, 1988).

2.3 Procedural Details

The experiment was computerized using the software z-Tree (Fischbacher, 1999). All of the interaction was anonymous. Most of the instructions were presented on the computer screen. At the very beginning, however, subjects were handed out a written overview that informed them about the work task and presented the basic structure of the experiment. Subjects were told that no aid was allowed for answering the problems (calculator, paper and pencil etc.) and that we would check this throughout the experiment. We ran twelve sessions, four sessions in each of the three treatments. A total of 240 subjects participated. We invited the same number of females and males in each session and ended up with 121 female and 119 male participants. A session lasted, on average about 90 minutes. Subjects were students from the University of Bonn. Ten points in the experiment were exchanged for 0.17 Euro (1 Euro $\sim$ 1.20 US Dollar). Average earnings were 20.80 Euro.

3 Results

In this section we present the main results. In section 3.1, we start by investigating whether subjects who opt for a variable pay contract produce more than subjects who

[5] The exact wording of the question (translated from German) is as follows: How do you see yourself: "Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks? Please tick a box on the scale, where the value 0 means: 'unwilling to take risks' and the value 10 means: 'fully prepared to take risk'."

prefer to work for a fixed payment. In section 3.2 we focus on the role of sorting. First, we assess the role of productivity as an influencing factor for the selection into an incentive system. We then explore how other worker attributes — including risk preferences, relative self-assessment, overconfidence, social preferences, gender, and personality — determine the sorting decision. Finally, in section 3.3, we study how effort choices respond to the different incentive schemes.

3.1 Output

Our first result concerns output differences between variable and fixed payment schemes. We expect a positive output effect of variable pay schemes for two reasons. First, more productive subjects are likely to self-select into variable pay schemes, as we will address in more detail in the next section. Second, incentive theory predicts that subjects should work at least as hard in the variable pay schemes as in the fixed payment. Our first result verifies this expectation.

Result 1. *Output in all variable pay schemes is higher than output under the fixed wage regime.*

Figures 1 and 2 provide evidence for Result 1. Charts (a), (b) and (c) of Figure 1 depict realized output during the 10-minute work period (step 8 of the experiment) in the three different treatments. The upper histogram of each chart reflects the number of correctly solved problems for subjects who have self-selected into the fixed payment scheme while the lower shows the output distribution for subjects who opted for variable pay.

The figure clearly confirms that subjects on variable pay produce much more than those who work for a fixed payment. In all treatments the output distribution in the variable pay condition is shifted to the right compared to the respective output distribution in the fixed payment condition. The hypothesis that the fixed and variable output distributions are the same is rejected by a Wilcoxon rank-sum test at any conventional level (p-value < 0.0001) in each treatment. Subjects with a piece-rate contract solved on average 59.17 problems compared to 31.50 problems solved by subjects who worked for the fixed payment in the same treatment. The respective numbers in the tournament treatment are 61.03 versus 32.92 and in the revenue-sharing treatment 55.47 versus 34.48. OLS regressions (not reported here) of individual output on an indicator variable for variable pay and a constant substantiate that these output differences between the two self-selected groups of subjects are statistically significant at any conventional level in every treatment.

Figure 2 restates this result in a different way, which is also informative on another dimension. The horizontal bars in the figure represent how much time (in seconds) subjects with a particular remuneration contract need on average to enter the correct solution to a problem with a certain degree of difficulty. The brighter the bars the more difficult is the respective problem. For example, in the piece-rate treatment subjects who work under the fixed payment scheme need on average about 25 seconds to correctly answer a problem of difficulty level 5. Those who work on a piece-rate contract, however, need only about 13 seconds. The figure illustrates that regardless of the treatment, subjects in the variable pay schemes solve problems much faster than those working for a fixed payment. This holds for problems of all difficulty levels but is most pronounced for relatively tough problems. This pattern is partly explained by the fact that the error rate, which generally rises with the degree of difficulty, is higher for subjects in the fixed payment schemes.[6]

3.2 Sorting

The output differences observed in the previous section are most likely the result of sorting and different effort responses. In order to explain the output differences and to understand the nature of the sorting process, we therefore start this section by studying the role of individual productivity for the sorting decision. In a second step we investigate how sorting is affected by other personal attributes.

3.2.1 Productivity

In order to guide our discussion on productivity sorting, we allude to a simple model that sheds light on how productivity is expected to affect the choice between the fixed-payment contract w_F defined in equation (1) and a variable remuneration contract as defined in equations (2) to (4).

Suppose that subjects are endowed with a utility function, $u(w,e) = u(w - c(e))$, that depends positively on the wage, w, and negatively on the level of effort, e, due to effort costs, $c(e)$ with $c^e > 0$. Assume that individual output, x_i, depends on subject i's ability, θ_i, and effort, $e_i \geq 0$, according to the production function $x_i = \pi(\theta_i, e_i) + \varepsilon_i$ where $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$, π^θ, $\pi^e > 0$ and $\pi^{e\theta} \geq 0$. Subjects are heterogeneous with respect to ability, which we assume to be continuously distributed on the interval $[\underline{\theta}, \overline{\theta}]$ according

[6] Subjects solved 95 percent of problems on their first attempt when the degree of difficulty was equal to 1, 92 percent when the degree of difficulty was 2, 88 percent when it was 3, 82 percent when the degree of difficulty was 4 and 78 percent when it was 5. Holding constant the degree of difficulty a Probit analysis shows that subjects who selected the fixed-payment contract are 3 percent more likely on average to enter a wrong answer.

to the cumulative distribution function $F(\theta)$.

If utility maximizing risk-neutral subjects choose between the fixed-payment contract w_F (equation (1)) and the piece-rate contract w_{PR} (equation (2)), it is straightforward to show that subjects whose productivity exceeds a certain threshold value optimally opt for the piece-rate contract, while subjects with lower productivity prefer the fixed-payment contract. This productivity threshold, $\hat{\pi}$, is given by $\hat{\pi}(\theta, e^*) = \frac{\alpha + c(e^*) - c(e^{min})}{\beta}$, where e^{min} and e^* denote the effort choices that maximize utility in the fixed payment regime and in the piece-rate regime, respectively. The term $c(e^*) - c(e^{min})$ captures the disutility that results when effort is raised from e^{min} to e^*.[7] In line with intuition, the productivity threshold increases in the level of the fixed payment α, and it decreases in the attractiveness of the piece rate β.

In the tournament, productivity sorting is plausible but not as obvious as in the piece-rate treatment. The reason is essentially the strategic nature of tournaments. According to our simple framework, a risk-neutral subject optimally participates in the tournament if $\gamma \cdot Prob\{\pi_i(\theta_i, e_i^*) - \pi_j(\theta_j, e_j^*) > \epsilon_j - \epsilon_i\} \geq \alpha - c(e^{min}) + c(e_i^*)$. Thus the sorting decision does not only depend on own productivity and luck but also on the expected productivity of the other player who has sorted into the tournament. Therefore the existence of a unique sorting threshold depends on various distributional assumptions and is not guaranteed. For example, if luck is absent, i.e., $\sigma_\epsilon = 0$ in the production function, and ability is continuously distributed on a closed interval $[\underline{\theta}, \overline{\theta}]$, a more able contestant has an optimal effort response function that ensures winning the tournament against a less able competitor.[8] Since the most able subject always wins — and consequently enters the tournament — it is not optimal for a less productive person to compete. In this setting no tournament takes place, as only the most productive individual optimally opts for tournament incentives.[9] On the other hand, everybody will participate in the tournament if luck is sufficiently important relative to productivity differences. Finally, a sorting equilibrium, in which subjects whose ability exceeds a threshold $\hat{\theta}$ with $\underline{\theta} < \hat{\theta} < \overline{\theta}$ sort into the tournament and less able subjects select into the fixed payment scheme, may exist for intermediate cases. Productivity differences are likely to dominate luck in determining output and thus

[7] Note that, in our experiment, e^{min} captures the cost of remaining in the lab, sitting silently in front of the computer during the 10-minute work period.

[8] Since it is never more costly for the more able contestant to provide the same level of effort as a less productive a competitor, he can always provide the same level of effort as a competitor and thereby ensure winning the tournament.

[9] If the best subject participates, it is not profitable for a non-participant to deviate from the non-participation strategy, since the tournament against the best subject would be lost with certainty. Entering the tournament is a weakly dominant strategy for the most able subject as he receives the outside option when no tournament takes place.

the likelihood of winning in our experiment. Therefore, it is quite reasonable to expect an outcome in which more productive workers are more likely to participate in the tournament than less productive workers.

As in the tournament treatment, the sorting prediction in the revenue-sharing treatment depends on the assumed distributions. It is possible in theory to fix parameters such that either all subjects are expected to join the team, or no one is expected to join the team, or some are, either with or without a unique threshold. Note, however, that highly productive types, whose productivity exceeds $\frac{2(\alpha - c(e^{min}) + c(e^{*}))}{\beta}$, should always sort into the revenue-sharing scheme. This is because they can attain higher utility than under fixed wages even if their team partner does not produce anything. Abstracting from effort costs the corresponding critical output is 80 correct answers during the 10 minute working time. We therefore expect average productivity to be higher among team participants than among subjects in the fixed wage scheme. In sum, productivity sorting is likely to occur in all treatments, and especially in the piece rate treatment. Our second result confirms these conjectures.

Result 2. *In all treatments there is systematic productivity sorting. On average, the more productive a worker, the more likely he self-selects into the variable pay scheme.*

Support for Result 2 comes from Figures 3 and 4 as well as from Tables 1 and 2. Figure 3 contains three charts, each of which compares the distributions of productivity (measured by Productivity Indicator 3) of subjects who sorted into the fixed payment scheme (upper histogram) and of subjects who sorted into the variable payment scheme (lower histogram) in a particular treatment. The fractions of subjects who self-select into the variable pay are 60.0 percent in the piece-rate, 48.75 in the tournament and 58.75 percent in the revenue-sharing treatment. Chart (a) of Figure 3 clearly confirms that those workers who self-select into the piece rate are more productive. Charts (b) and (c) show the same finding for the tournament and the revenue-sharing treatments, respectively. In line with our discussion above, the productivity histograms for subjects in the revenue-sharing treatment also reveal that all subjects whose productivity exceeds 40 answers in 5 minutes, and who are probably expecting to produce more than 80 corrects answers in the 10 minute work period, sort into the revenue-sharing scheme.

Wilcoxon rank-sum tests verify that the differences shown in Figure 3 are statistically significant in all treatments (p-values < 0.0001). Moreover, the differences in mean productivity are quite sizeable. In the piece-rate treatment subjects who later opt for the piece-rate contract have an average productivity (measured by Productivity Indicator 3) of 26.7 correct answers per five minutes compared to an average productivity of 14.2 of subjects who sort into a fixed-payment contract. Corresponding numbers are 25.8 versus 14.8 for the tournament treatment, and 24.5 versus 14.6 for the revenue-sharing treatment.

The result that more productive workers are more likely to sort into the variable payment holds regardless of which of the three productivity indicators is used.[10] In the upper part of Table 1 we report medians of all three productivity indicators, elicited in steps 1 to 3 of the experiment. The table shows, for example, that the median time needed to solve a problem amounts to 8 seconds for subjects who opt for the piece-rate contract and to 28.5 seconds for those who opted for the fixed payment in the piece-rate treatment. When paid for speed, subjects get faster in general (see Productivity Indicator 2), but the substantial productivity differences between the groups remain (7 seconds vs. 20.5 seconds). A similar pattern is observed in the tournament treatment and in the revenue-sharing treatment. Overall the productivity differences are highly significant in all three treatments as shown by p-values of median tests reported in Table 1, one exception being the median difference for Productivity Indicator 1 in the revenue-sharing treatment.[11]

The significance of productivity sorting is further substantiated by Probit regressions, in which the latent variable is the propensity to opt for the variable pay alternative (i.e., depending on the treatment, piece rate, tournament, or revenue sharing). Not surprisingly, the coefficient estimates reported in Table 2 show unequivocally that more productive subjects are significantly more likely to sort into the variable pay schemes than into the fixed payment scheme. In Table 2 (as well as in all other regression tables in this paper) we display the coefficients, the standard deviations and the marginal effects. The negative signs of the coefficient estimates for Productivity Indicator 1 and 2 indicate that the faster a subject solves the respective problems in steps 1 and 2 of our experiment, the more likely he is to opt for the variable pay contract. The estimated marginal effect of -0.031 for Productivity Indicator 2 in the piece-rate treatment indicates that taking one second longer to correctly answer the question reduces the probability to choose the piece rate by 3.1 percent. For subjects whose observations are censored, we construct a dummy

[10] All three productivity measures are also significantly correlated with the last math grade in high school — a measure that ranges from 1, the worst grade, to 15, the highest grade — and with the grade in the Abitur — a measure that ranges from 4.0, the worst grade, to 1.0, the best grade. The Spearman rank correlations and corresponding p-values of math grades and Productivity Indicators 1 to 3 are respectively: -0.28 (p-value < 0.001), -0.20 (p-value < 0.008), and 0.28 (p-value < 0.001). The Spearman rank correlations and corresponding p-values of Abitur grades and Productivity Indicators 1 to 3 are respectively: 0.29 (p-value < 0.001), 0.21 (p-value < 0.006), and -0.20 (p-value < 0.003). Our verbal IQ-measure is also positively correlated with productivity. However, the correlation is weaker and statistically significant at the 10 percent level only for Productivity Indicator 3.

[11] The p-values correspond to continuity corrected Pearson $\chi^2(1)$ statistics.

variable, which takes a value of one if a subject did not answer the question within 30 seconds. The resulting estimate in the piece-rate treatment reveals that theses subjects have a 57.5 percent lower probability to enter the piece rate.

The positive coefficients for Productivity Indicator 3 reveal that subjects in all treatments are more likely to select into a variable compensation system the more problems they solved in step 3 of the experiment. The estimated marginal effect (reported in parenthesis) of Productivity Indicator 3 in the piece-rate treatment (Column (3)) implies that answering one additional question in the 5 minute work period makes a subject 3.9 percent more likely to sort into the piece-rate scheme. It turns out that the marginal effect is strongest and the fit of the model is best in the piece-rate treatment, indicating that sorting leads to the most clear-cut partition of the productivity distribution in the piece-rate treatment. This is plausible given that sorting in the piece rate treatment does not depend on strategic considerations and beliefs and is therefore considerably less complex than sorting in the tournament and revenue-sharing treatments.

An important implication of productivity sorting is that the average productivity of a selected group depends on the relative attractiveness of the given contracts. Recall from our discussion above that the theoretical productivity threshold in the piece-rate treatment increases in the level of the fixed payment α. Consequently, we would expect fewer and more productive workers to select into the piece-rate scheme when the fixed payment alternative becomes more attractive. Similarly, more productive workers should choose the tournament or the revenue-sharing scheme as the the level of the fixed payment increases. These predictions are born out by our data on hypothetical sorting decisions elicited in step 7 of the experiment. The correlation between individual productivity and the threshold α, i.e., the lowest fixed wage a subject just prefers over the variable payment, is positive and highly significant (p-values of Spearman rank correlations < 0.001 in all treatments).

Figure 4 shows the sorting pattern in all treatments for different (hypothetical) fixed payments using Productivity Indicator 3. Panel (a) displays the results for the piece-rate treatment. The bars in the lower part of the panel reveal that the fraction of workers who self-select into the piece rate is the higher the lower the fixed payment, which is displayed in steps of 50 points on the horizontal axis. For example, when the fixed payment is 50 or 100 points, all workers prefer the piece rate, while 60 percent prefer the piece rate when the fixed payment is 400 points, the level actually implemented in the experiment. If the fixed payment is as high as 800 points, almost nobody selects into the piece-rate scheme anymore.

The consequences for average productivity of the selected groups are displayed in the top panel. Dark dots represent subjects sorting into the piece rate and grey diamonds represent subjects sorting into the fixed payment. The dashed grey horizontal line reflects average productivity of all subjects who participated in the piece-rate treatment. Since for very low fixed wages all workers prefer the piece rate to fixed payments, the average productivity in the piece-rate group coincides with the overall average productivity. As the fixed payment increases, typically the least productive workers from the piece-rate group start sorting into the fixed payment. This leads to an increase in the average productivity in the piece rate group and to a relatively low productivity level in the fixed wage group. As the level of the fixed payment increases, more productive workers select into the fixed payment group such that the average productivity in this group eventually approaches the overall average.

The sorting pattern is similar in all three treatments as shown in panel (b) for the tournament and in panel (c) for the revenue-sharing treatment. As the fixed wage becomes more attractive, fewer and fewer subjects self-select into variable pay. Those workers who switch to the fixed payment as a response to an increased fixed payment are typically among the least productive of the subjects on the variable payment scheme. This leads to the increase in productivity of the variable payment group.

3.2.2 Relative Self-Assessment and Overconfidence

A fundamental difference between piece rates on the one hand and tournaments or revenue-sharing on the other hand is that in the former scheme payoffs depend only on one's own performance and are independent of other workers' outputs. As a consequence, beliefs about other workers' productivity are irrelevant for the sorting decision in the piece-rate treatment but could affect the sorting decision in both the tournament and the revenue-sharing treatment. We therefore expect that beliefs about their relative rank should affect the sorting decision in these treatments while no such effect is expected in the piece-rate treatment. The next result partly confirms this expectation.

Result 3. *Relative self-assessment affects the decision to select into a tournament.*

Result 3 is supported by the estimation results shown in Table 3. We estimate the effect of a subject's relative self-assessed rank, elicited in step 5 of the experiment, on the propensity to opt for the variable pay contract in each treatment. It turns out that relative self-assessment significantly predicts sorting into the variable pay condition in all three treatments (see Columns (1), (3) and (5) of Table 3). The significant negative coefficient estimate for the variable "relative self-assessment" means that subjects are more likely to select into the variable pay schemes the more productive they believe they are relative to the other participants. Note, however, that this finding just reflects the productivity sorting result we have shown in the previous section if relative self-assessed ranks and true ranks are highly correlated, which is in fact true. Participants have a relatively good notion of their productivity relative to that of others. The correlation between a person's self-assessed rank and his/her true rank based on Productivity Indicator 3 is 0.69.[12] The relevant question, therefore, is whether relative self-assessment predicts the sorting decision even after controlling for productivity. Once we control for productivity, relative self-assessment predicts sorting only in the tournament treatment, but not in the piece-rate and the revenue-sharing treatment (see Columns (2), (4) and (6) of Table 3).[13] The marginal effect estimate of relative self-assessment in the tournament treatment is sizable: A subject with a more positive self-assessment of only one rank is 3 percent more likely to enter the tournament than a less optimistic but equally productive subject. In light of Result 3, we expect that overconfidence about one's own rank should affect the sorting decision in the tournament treatment, but not in the other two treatments. Our findings on overconfidence are summarized in the following result:

Result 4. *On average, the prevalence of overconfidence is not particularly pronounced. However, overconfident subjects are more likely to self-select into tournaments.*

We measure overconfidence as the difference between a subject's self-assessed rank and true rank. Among all subjects, 48 percent overestimate their rank, and 36 percent underestimate their rank. Table 4 provides estimates of the impact of overconfidence on the decision to sort into the three different variable pay schemes, controlling for productivity. Overconfidence is measured by the variable "relative overassessment." A positive value indicates by how many ranks a subject overestimates his true rank. Since this measure limits the extent to which a person can be overconfident or underconfident, we weigh observations by the probability that a subject can plausibly have an overassessment or underassessment at all. Subjects in median ranks, i.e., subjects in the 45th to 55th percentile of the productivity distribution receive a weight of one, subjects in the 40th to 45th percentile and subjects in the 55th to 60th percentile receive a relative weight of 8/9, subjects in the 35th to 40th percentile and subjects in the 60th to 65th percentile

[12] This correlation is somewhat stronger in the piece-rate treatment (0.74) and in the tournament treatment (0.72) than in the revenue-sharing treatment (0.63).

[13] Note that the standard errors of coefficient estimates in Table 3 are large due to the strong correlation of Productivity Indicator 3 and relative self-assessment.

receive a relative weight of 7/9, and so forth. Observations of subjects in the lowest ranks of the productivity distribution receive a weight of zero since they can never bee too pessimistic about their own rank. Likewise, observations of top-ranked subjects are assigned a weight of zero since they can never be too optimistic about their rank. The positive and statistically significant coefficient estimate for relative overassessment indicates that overconfident subjects are more likely to participate in a tournament. However, overconfident subjects are not significantly more likely to select into piece-rate or revenue-sharing schemes.[14]

3.2.3 Risk Preferences

A potentially very important personal characteristic that may affect the sorting decision into different incentive schemes is an individual's attitude towards risk, not least because contract theory regards risk as an important constraint to offering incentives.[15]. To illustrate the role of risk preferences in our context, assume that risk preferences are captured by the curvature of the utility function $u(w, e)$. If subjects have the choice between the fixed wage contract w_F and the piece-rate contract w_{PR}, the productivity threshold for individual i is now determined by the comparison of expected maximum utility in both regimes. Since there is no earnings risk in the fixed payment scheme the expected utility derived from the fixed payment contract equals the utility of the fixed payment. But the expected utility of the piece-rate contract is smaller than the utility of expected piece-rate earnings ($Eu^*_{PR} < u\beta\pi(\theta, e^*) - c(e^*)$) for risk averse subjects, and larger for risk loving subjects , ($Eu^*_{PR} > u\beta\pi(\theta, e^*) - c(e^*)$). Therefore, the productivity threshold at which the piece-rate contract becomes the preferred choice is higher for risk averse subjects and lower for risk loving subjects. Analogous reasoning applies if subjects have the choice between the fixed-payment contract and the tournament or the revenue-sharing contract. Therefore, we expect risk preferences to affect the sorting decision in all three treatments. However, earnings uncertainty is highest in the tournament treatment because earnings are either quite high or zero. In addition, tournament participants face uncertainty about

[14] These results are robust to alternative weighting schemes. For example, we have estimated a model as in Table 4 but using only information on workers in the 5th to 95th percentile of the true productivity distribution. We have also sequentially trimmed the sample even further using only information on workers in the 10th to 90th percentile, 15th to 85th percentile, 20th to 80th percentile, and so forth. No matter which weighting scheme we use, we always find that overconfident workers are significantly more likely to sort into tournaments while no significant pattern arises for the piece-rate or the revenue-sharing treatment.

[15] Principal-agent theory has emphasized that risk-averse workers dislike the income risk that is associated with variable pay when output depends upon factors beyond their control, which triggers a trade-off of risk and incentives (see Prendergast, 1999 and references therein)

the contestant's ability, such that we expect the impact of risk preferences on sorting to be strongest in the tournament treatment.

Result 5. *Risk preferences affect the sorting decision. Risk averse workers are less likely to self-select into tournaments and piece rates.*

Estimates of the impact of risk preferences on the decision to sort into the different incentive systems in Table 5 support this result. In the reported Probit regression models we measure risk preferences by subjects' responses to the risk question elicited in step 12. We prefer this risk measure over the lottery measure elicited in step 11 since several subjects did not have a unique switching point. It is not clear how these observations should be treated. Moreover, we are quite confident that the answers to the risk question are a good measure of risk preferences. First, Dohmen et al. (2005) have shown that it predicts lottery choices in a paid field experiment. Second, we find a strong correlation between subjects' answers to the risk question and the lottery choices in our experiment.

The coefficient estimates imply that, controlling for productivity, risk preferences significantly affect the sorting decision in the piece-rate treatment and in the tournament treatment. The coefficient estimate for the effect of risk attitude on the decision to sort into the piece-rate contract in Column (1), for example, indicates that a one point higher indication of willingness to take risk on the eleven-point scale makes a subject 5.8 percent more likely to opt for the piece-rate contract for a given level of productivity. The estimates in Columns (1) to (3) also indicate that risk preferences matter most, both in terms of quantitative importance and in terms of statistical significance, in the tournament treatment. This probably reflects the fact that uncertainty is greatest in the tournament treatment.

In the piece-rate treatment, we would expect risk preferences to matter most for workers who are close to the productivity threshold that is relevant for a risk neutral agent. After all, very productive subjects who are far above this productivity threshold should always prefer the piece rate regardless of their risk preferences. Likewise, very unproductive subjects should always prefer the fixed payment. In order to test this implication, we estimated the model from Column (1) on the sample of 53 subjects who solved between 10 and 30 problems in step 3 of the experiment.[16] The results support our prediction: the estimated marginal effect of our risk measure is larger in this sub-sample (equal to 9.4

[16] We chose these numbers because 20 problems would be a risk-neutral subject's productivity threshold in the piece-rate treatment if the following is true: the disutility that results when effort is raised from minimum effort level in the fixed payment to optimal effort in the piece treatment is negligibly small, and subjects expect to produce twice as much when the working time is twice as long.

percent) and statistically significant at the 5 percent level. Hence, risk preferences matter more for subjects whose productivity is closer to the productivity margin.

3.2.4 Social Preferences

Traditional contract theory is based on the assumption that principals and agents are solely interested in their own material payoffs. In contrast, there is by now considerable evidence indicating that a substantial fraction of people also care about reciprocal fairness (see the overviews by Camerer, 2003; Fehr and Gächter, 2006; and Fehr and Schmidt, 2000). The co-existence of selfish and reciprocally motivated agents changes the optimality conditions of different types of contracts. For example, Fehr, Klein and Schmidt (2005) find in their experiment that contracts, which are optimal when all actors are selfish, may be less efficient when there is a minority of people who care about fairness. Furthermore, contracts that are inefficient if all actors are selfish may achieve surprisingly high levels of efficiency when there are some fair-minded people. Theoretical implications of social preferences for optimal contracting are derived in Grund and Sliwka (2005) and Englmaier and Wambach (2005). Given the relevance of social preferences for optimal contracting it is important to understand whether they also affect the sorting of agents. To shed light on this issue we report in this section how subjects with given reciprocal preferences select themselves into different pay schemes.

Remember from section 2.2 that all subjects participated in a trust game. Since we made use of the strategy method, we know for each agent and for each transfer how much he is willing to pay back. In order to classify the agents with respect to their reciprocal inclination, we first determined for each subject the relation between transfer and back transfer. We ran simple OLS regressions of the back transfers on received transfers, forcing the slope to go through the origin. In a second step we grouped all subjects according to their individual back transfer slope. Subjects with a slope equal to zero are called *selfish* because they send back nothing irrespective of the first mover's transfer. 12.5 percent of the subjects are selfish. Subjects with a slope larger than zero but smaller than one are classified as *weakly reciprocal*. They pay back something but on average they pay back less than they have gotten from the first mover (22.9 percent). Finally, we call all subjects with a slope larger than or equal to one *reciprocal*. These subjects return at least as much as they have received from their first movers (64.6 percent).

In columns (1) to (3) of Table 8 we show Probit estimates of how reciprocal preferences affect the willingness to work under the three different variable pay schemes. The indicator variables classify subjects as either selfish, reciprocal or weakly reciprocal, where the latter is the reference category.[17] Controlling for productivity column (1) shows that social preferences play no role for the sorting decision between fixed wages and piece rates. Neither the variable selfish nor the variable reciprocal are close to being significant. This is not surprising as both, fixed wages and piece rates are individual pay schemes that provide no sensible basis for any sort of social comparison or fairness judgement. After all, the decision to work under fixed wages or piece rates does not affect payoffs of anybody else.

In contrast, tournaments seem to attract significantly less reciprocal subjects than fixed wage contracts, as Column (2) of Table 8 shows. One potential explanation is that people endowed with reciprocal preferences dislike interacting in competitive and non-cooperative environments where incentives are such that higher work effort produces a negative externality on others. Moreover, tournaments lead to extremely unequal outcomes. While the winner earns a lot, the loser gets nothing. If reciprocal subjects are averse to unequal outcomes, they may be willing to trade off lower expected pay with less inequity (see Fehr and Schmidt, 1999; and Falk and Fischbacher, 2006 for models that are compatible with this interpretation). No clear picture emerges from the revenue-sharing treatment. One reason may be that while payoffs are identical in the team, reciprocal workers may nevertheless be hesitant to join as they are afraid of being exploited. This would be the case if they cooperate and work hard, while the teammate free rides.

Recall that we also collected information about first movers' willingness to trust in the trust game. We can therefore also check whether different levels of trust affect worker self-selection. As a measure of trust we simply use the amount transferred in the trust game. The relative frequencies of sending 0, 20, 40, 60, 80, 100, and 120 points are 10.83 percent, 16.25 percent, 18.33 percent, 22.08 percent, 10.83 percent, 4.58 percent, and 17.08 percent, respectively. Columns (4) to (6) show whether willingness to trust predicts sorting decisions, controlling for productivity. It turns out that trust does not significantly explain any sorting decision. Our results on social preferences can therefore be summarized as follows:

Result 6. *Reciprocal subjects are less likely to enter tournaments than selfish subjects. Willingness to trust is no significant predictor of the sorting decision.*

[17] Note that the results are virtually the same if we use a binary measure, grouping selfish and weakly reciprocal workers in one group and reciprocal workers in the other group.

3.2.5 Gender and Personality

In this section we are interested in gender and personality as factors that may affect the sorting decision into variable pay. We think that investigating the impact of gender and personality on contractual choice is particularly interesting since it offers a potential explanation for the gender wage gap. If women are more likely than men to prefer non-competitive and non-variable pay, this would translate directly into lower average wages for women than for men. To motivate this intuition with non experimental data, we use data from the German Socio Economic Panel (GSOEP, 2004 wave), a large and representative panel data set. It turns out that in Germany women are much more likely to work in the public sector compared to men. The public sector is characterized by fixed wages and low risks concerning income variability and unemployment, but also by relatively low wages compared to the private sector. Considering full time employment, 32.96 percent of all women work in the public sector and 67.04 percent in the private sector. The respective numbers for men are 21.25 percent and 78.75 percent.[18] While this evidence is suggestive it could also be driven by different working conditions or employment protection laws. Our controlled experimental environment rules out these potential confounds and therefore allows a more direct assessment of gender sorting.

A first indication that women prefer fixed wages to a higher extent than men in our experiment is given by Table 7. It reports the relative numbers of female and male participants who self-select into variable pay or fixed wages. For example, while 74 percent of the male participants in the piece-rate treatment choose to work under the variable pay wage regime, only 45 percent of the female do so. Similar numbers prevail for the decision to work under tournament or revenue-sharing incentives. In Column (4) we pool the three treatments. 68 percent of the 119 male subjects prefer variable pay compared to only 44 percent of the 121 female subjects.

Of course, results in Table 7 do not correct for potential productivity differences between male and female participants. However, as Figure 5 shows, the gender differences are strong and robust if we compare only subjects with similar productivity levels. The figure shows for each treatment separately as well as for all treatments pooled the fraction of males and females in a given productivity cluster who sort into variable pay. We use the following clusters according to Productivity Indicator 3: less than 15, 15 to 19, 20 to 25, and above 25 problems solved. For example, in the piece-rate treatment, displayed in

[18] Numbers (in percent) for part time employment are: Women: 36.12 public, 67.04 private; Men: 28.17 public, 71.83 private. The higher percentages of people working part time in the public sector simply reflects the fact that it is easier to find part time jobs in the public sector than in the private sector.

the upper left panel, about 60 percent of the male participants who solved 15 to 19 correct answers in the 5 minute work period choose the variable pay, while only about 40 percent of the female participants with the same productivity level do. If we pool all treatments, shown in the lower right panel of Figure 5, we find that men are more likely to choose variable pay than women in each of the four productivity brackets.

The significance of our findings is shown in Table 8. This table reports Probit estimates for the choice of the variable pay contract vs. the fixed-payment contract. We pool observations from all three treatments. The dependent variable is 1 if a subject chooses the variable pay and 0 otherwise. We show the coefficients, the standard errors and the marginal effects. In the first column, the only regressor is a gender dummy which takes the value 1 if the subject is female and zero otherwise. The negative coefficient shows that women are less likely to sort into the variable pay schemes than men. The marginal effect indicates that, on average, female subjects are about 24 percent less likely to enter a variable pay scheme than male subjects. In the second column, we control for productivity. Not surprisingly, the respective coefficient is positive and highly significant. While the gender coefficient gets considerably smaller – with a marginal effect of about 15 percent – it still has the negative sign and remains significant at the 5-percent level. Thus differences in productivity cannot fully explain the different choices of women and men.

But what about differences in risk preferences? Many studies have shown that, on average, women tend to be more risk averse than men (see Croson and Gneezy, 2004 for an overview). These differences also hold true in our sample. Both behavior in the lottery experiment as well as responses to the risk question, reveal that women are more risk averse than men. Average responses to the general risk question are 4.83 for women and 5.96 for men (medians are 5 and 6, respectively).[19] This difference is highly significant (Wilcoxon rank-sum test $p<0.0001$). Given the importance of risk preferences for the sorting decision, shown above, the difference in risk preferences is a promising candidate for explaining gender differences. In fact, the third column in Table 8 shows that, as before, risk and productivity matter, but the gender coefficient becomes smaller and insignificant. The marginal effect is still negative but reduced to about 7 percent. Thus it seems that most of the gender differences can be attributed to differences in productivity and risk preferences. These findings are in some contrast to a recent paper by Niederle and Vesterlund (2005). In their study gender differences in sorting cannot be explained by either productivity or risk. Note, however, that in their set-up the decision is between piece rates and tournaments and not between fixed wages and variable pay as in our study. Since both options in their

[19] This is in line with the findings of Dohmen et al. (2005).

study are variable pay schemes, they are both associated with some risk and uncertainty. Therefore risk preferences may be less important to motivate the sorting decision than in our set-up. Regardless of why we observe the gender differences, however, both studies show an important interaction between gender and contractual choice, which is of relevance both from a general research as well as from a policy oriented perspective. We summarize our findings about the impact of gender on sorting into incentive schemes as follows:

Result 7. *Women are less likely to sort into variable payment schemes than men. This effect is at least in part driven by gender specific risk preferences.*

In the remaining part of this section we study the impact of personality. Little or nothing is known empirically about the relation between personality and sorting into different payment schemes or firms. This is surprising insofar as personality may matter to firms as much as other more standard personal characteristics. In fact, employer surveys suggest that so-called "soft skills" such as reliability or positive work attitudes are rated by employers as more important than prior work experience or technical skills (Regenstein, Meyer, and Hicks, 1998; Becci et al., 2005; Atkinson and Williams, 2003). But if these soft skills are important and if people with particular soft skills and personalities feel systematically more or less attracted to work in particular organizations or under particular pay schemes, the firm's decision about pay schemes should take this sorting into account.[20]

In our study we measure personality with the help of the 16 PA test, which was developed by Hermann Brandstätter and is described in Brandstätter (1988). The 16 PA test is a short form of the German-language version of Cattell's sixteen personality factor questionnaire (16 PF), an internationally well established personality assessment.[21] Our main result can be stated as follows:

[20] Firms actually do use personality tests in their hiring process, see Autor and Scarborough (2005).

[21] The German-language version of the 16 PF was developed by Schneewind, Schröder and Cattell (1983) and contains 192 items that compass sixteen primary scales of personality. The sixteen primary scales produce five independent categories (the so-called "Big Five") that describe individuals' personality at a broad level of abstraction. These five categories, or secondary factors, are commonly labelled as high or low conscientiousness (Q_I), neuroticism vs. emotional stability (Q_{II}), independence (Q_{III}), tough-mindedness (Q_{IV}) and extroversion (Q_V). In order to obtain measures of these five secondary factors from the 16 PA test, which we administered, we follow the procedure described by Brandstätter (1988). Subjects are presented 32 conflictive adjective pairs, which describe traits. For each adjective pair, subjects indicate on a 9-point scale which adjective is more in line with their personality. This provides us with 32 measures. Linear combinations of these 32 measures produce the five secondary factors Q_I – Q_V. The coefficients of these linear combinations were determined in a regression analysis on a sample of 300 individuals who had completed both the 16 PF test and the 16 PA test. Each of the five secondary factors, which were derived from the 16 PF following the procedure described in Schneewind, Schröder and Cattell (1983), was regressed on all 32 measures that the 16 PA test produced. The estimated coefficient vector provides the parameters for the linear combinations that map out the five secondary factors. Since the regressions were estimated separately for men and women, we obtain different coefficient vectors for men and women.

Result 8. *Personality affects sorting into different payment schemes. The sorting patterns are different for different incentive schemes as well as for men and women.*

Support for Result 8 comes from Table 9, which reports estimates from simple Probit regressions, in which the dependent variable takes value 1 if a subject selected the variable pay and zero otherwise. For each of our three treatments the regressors are the five personality categories. Since the weights are quite different for women and men, we report separate regressions for female and male subjects. We find that several of the categories explain the sorting decision into variable pay. Let us first consider male participants. For men, emotional stability is positively related with the decision to choose the piece rate. Conscientiousness or norm orientation turns out to be significantly negative for men who opt for the tournament. In light of the statistically most important adjectives behind this category we can interpret this finding as indicating that these subjects are open for change, reckless, and good in cogitation. None of the five categories explains the sorting decision into revenue-sharing teams. Turning to female subjects we find that extroversion is negatively related to the decision to work under piece-rates. Interpreting this in terms of the attributes with largest coefficients in the linear combination of the factor, these women are businesslike, headstrong, adventurous, and not warmhearted. Emotional stability is positively associated with tournament participation. Thus women who are self-confident, reckless and who can rather easily deal with defeat prefer tournaments. Finally, the less pronounced independence (or dominance) is, the more likely women sort into revenue-sharing teams. Interestingly these are women who indicate that they are rather shy instead of self-confidence, the opposite to tournaments. These women also consider themselves as mentally stable and rather unwilling to experiment.

Summing up we find that personality matters for the sorting decision and it matters in a different way for men and women. For example, while males with lower norm-orientation are more likely to join tournaments, this incentive scheme is preferred by women who are rather tough. In terms of our general research question the data suggest that different incentive devices are quite likely to systematically attract people not only with different abilities, preferences and self assessments but also with different personalities.

3.3 Effort Provision

In this final results section we briefly discuss how different incentive schemes affect effort provision. It is intuitively clear that subjects in the variable pay schemes should provide at least as much effort as subjects with the fixed-payment contract. This follows from the fact that all variable payment schemes add an explicit reward for providing effort. Consequently, we also expect that subjects in variable pay schemes feel more stressed and get more exhausted than subjects who work for the fixed payment. These expectations are all borne out by the data:

Result 9. *Subjects in variable pay schemes provide more effort than those who receive fixed payments. In addition they report more stress and exhaustion.*

Table 10 compares average self-reported effort levels, stress and exhaustion for two subgroups: subjects who sorted into the variable compensation scheme and subjects who opted for the fixed compensation scheme. Panel (a) shows the results for the piece-rate treatment, while panels (b) and (c) show the outcomes for the tournament and revenue-sharing treatments, respectively. Columns (1) to (3) of the table refer to the 5-minute work period (step 3), in which all subjects worked under the exact same incentives. For example, mean effort in the piece rate treatment is 5.63 for those who later selected into the piece rate while it is 5.50 for those who later prefer the fixed pay. This difference is not statistically significant as the corresponding p-value in Column (3) reveals. In fact, for all treatments there are no statistically significant differences in effort, stress, and exhaustion between the two subgroups.

Things change a lot, however, when subjects work in their preferred incentive scheme during the 10-minute work task (see Columns (4) to (6)). In the piece-rate treatment, e.g., mean effort is now 6.00 for workers receiving a piece rate, while it is only 4.25 for the fixed wage group. This difference is highly statistically significant as the p-value in Column (6) shows. Results in Table 10 reveal that in all treatments, subjects with a variable compensation contract provide significantly more effort and feel significantly more stressed than subjects in the fixed payment scheme. Regression estimates from an Ordered Probit model with effort measured on the 7-point scale as dependent variable reinforces the result that subjects working for variable pay provide more effort, even when controlling individual productivity.[22] A comparison of efforts in Columns (5) and (2) reveals that subjects who select into the fixed-payment contract put forth less effort than they previously did when they were working in the piece rate condition. Sign-rank tests, which are not reported in the table, confirm that this slacking off is statistically significant. Finally, Table 10 indicates that subjects who work for variable pay tend to get more exhausted, but differences in exhaustion levels are not significant in all treatments.

[22] The results are available upon request.

4 Concluding Remarks

In this paper we have provided controlled laboratory evidence on the importance of multi-dimensional sorting. Productive workers are more likely to self-select into variable payment schemes when offered a fixed payment scheme as an alternative. This productivity sorting explains a substantial part of output differences observed in variable versus fixed payment schemes. Controlling for productivity, workers are more likely to prefer a fixed payment scheme the more risk averse they are, especially when the choice is between tournaments and fixed wages. Tournament schemes not only attract more risk tolerant individuals, but also more overconfident and more selfish workers. Variable payment schemes attract fewer women, an effect that is substantially driven by an underlying gender difference in risk attitudes. Finally, personality systematically affects the sorting decision but differently for men and women. Besides their impact on sorting, incentives of course also affect effort provision. In our study workers provide more effort in pay for performance schemes than in to fixed payment schemes. Moreover, they report higher levels of stress and exhaustion.

Our findings on gender and risk complement recent evidence from survey data that suggest that risk attitude are important for occupational sorting: Bonin et al. (2006) find that individuals who are more willing to take risks are more likely to work in occupations with higher earnings variability. Dohmen *et al.* (2005), who observe that risk averse workers are more likely to be employed in the public sector, i.e., in jobs characterized by low earnings risk and low risk of job loss.

Multi-dimensional sorting has several important implications. For example, when thinking about optimal incentives, organizations should not only focus on effort effects but also consider the self-selection of different types of workers. Given that many of the discussed personal attributes, such as risk aversion or overconfidence are difficult to observe in the process of recruitment, an incentive scheme may also serve the purpose of a screening device. Of course, sorting is not only relevant between but also within firms. Firms can offer different career paths to get the right people on the right job. Our results also shed light on the question why firms use different incentive schemes even when operating in similar environments. A possible explanation is simply that they have different preferences regarding the composition of their workforce, which they manage to attract with different organizational features. Our findings on gender and risk attitudes point to a potential channel for gender differences in occupational choice, career choice, and ultimately for the existence of the gender wage gap.

References

Atkinson, J., and M. Williams (2003): "Employer Perspectives on the Recruitment, Retention and Advancement of Low-Pay, Low-Status Employees," Government Chief Social Researcher's Office, UK.

Autor, D. H., and D. Scarborough (2005): "Will Job Testing Harm Minority Workers? Evidence from the Retail Sector," MIT, Department of Economics.

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

Baron, D., and D. Besanko (1984): "Regulation and Information in a Continuing Relationship," *Information Economics and Policy*, 1, 447–470.

Becci, N., J. Hurstfield, L. Miller, R. Page, and K. Akroyd (2005): "What Employers Look for When Recruiting Unemployed and Inactive: Skills, Characteristics, and Qualifications," Department of Works and Pensions, Research Report No. 295.

Berg, J., J. Dickhaut, and K. McCabe (1995): "Trust, Reciprocity and Social History," *Games and Economic Behaviour*, 10, 122 – 142.

Bognanno, M. L. (2001): "Corporate Tournaments," *Journal of Labor Economics*, 19(2), 290–315.

Bohnet, I., and D. Kübler (2004): "Compensating the Cooperators: Is Sorting in the Prisoner's Dilemma Game Possible?," Working Paper.

Bonin, H., T. Dohmen, A. Falk, D. Huffman, and U. Sunde (2006): "Cross-sectional Earnings Risk and Occupational Sorting: The Role of Risk Attitudes," IZA Discussion Paper No. 1930.

Brandstätter, H. (1988): "Sechzehn Persönlichkeits-Adjektivskalen (16 PA) als Forschungsinstrument anstelle des 16 PF," *Zeitschrift für experimentelle und angewandte Psychologie*, 35(3), 370–391.

Brown, C. (1990): "Firms' Choice of Method of Pay," in *Do Compensation Policies Matter?*, ed. by R. G. Ehrenberg, pp. 165–182, Ithaca, New York. ILR Press, Cornell University.

——— (1992): "Wage Levels and Methods of Pay," *Rand Journal of Economics*, 23(3), 366–375.

——— (1994): "Pay and Performance," in *Labour Market and Economic Performance*, ed. by T. Tachibanaki, pp. 216–250, London, U.K. St. Martin's Press.

Bull, C., A. Schotter, and K. Weigelt (1987): "Tournaments and Piece Rates: An Experimental Study," *Journal of Political Economy*, 95(1), 1–33.

Cadsby, B. C., F. Song, and F. Tapon (2005): "The Effects of Compensation Schemes on Self-Selection and Productivity: An Experimental Investigation," Department of Economics, University of Guelph.

Camerer, C. F. (2003): *Behavioral Game Theory*. Princeton University Press, Princeton.

Camerer, C. F., and D. Lovallo (1999): "Overconfidence and Excess Entry: An Experimental Approach," *American Economic Review*, 1(89), 306–318.

Chiappori, P.-A., and B. Salanié (2003): "Testing Contract Theory: A Survey of Some Recent Work," in *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, ed. by M. Dewatripont, L. Hansen, and P. Turnovsky, New York. Cambridge University Press.

Croson, R., and U. Gneezy (2004): "Gender Differences in Preferences," unpublished manuscript.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2005): "Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally-Validated Survey," IZA Discussion Paper No. 1730.

Ehrenberg, R. G., and M. L. Bognanno (1990): "Do Tournaments Have Incentive Effects?," *Journal of Political Economy*, 98(6), 1307–1324.

Englmaier, F., and A. Wambach (2005): "Optimal Incentive Contracts under Inequity Aversion," IZA DP No. 1643.

Eriksson, T., S. Teyssier, and M.-C. Villeval (2005): "Does Self-Selection Improve the Efficiency of Tournaments?," GATE WP.

Eriksson, T., and M.-C. Villeval (2004): "Other-Regarding Preferences and Performance Pay. An Experiment on Incentives and Sorting," IZA Discussion Paper 1191.

Fahr, R., and B. Irlenbusch (2000): "Fairness as a Constraint on Trust in Reciprocity: Earned Property Rights in a Reciprocal Exchange Experiment," *Economics Letters*, 66(3), 275–282.

Falk, A., and E. Fehr (2002): "The Power and Limits of Tournaments," unpublished manuscript, University of Zurich.

Falk, A., and U. Fischbacher (2006): "A Theory of Reciprocity," *Games and Economic Behavior*, 54(2), 293–315.

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

Fehr, E., and S. Gächter (2006): "Reciprocity and Contract Enforcement," in *Handbook of Experimental Economic Results*, ed. by C. Plott, and V. Smith, Amsterdam: North-Holland. Elsevier Science.

Fehr, E., A. Klein, and K. Schmidt (2005): "Fairness and Contract Design," mimeo, University of Munich.

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

——— (2000): "Fairness, Incentives, and Contractual Choices," *European Economic Review*, 44(4-6), 1057–1068.

Fernie, S., and D. Metcalf (1999): "It's Not What You Pay, It's the Way that You Pay It.," *Labour*, 13(2).

Fischbacher, U. (1999): "Z-tree. Zurich Toolbox for Readymade Economic Experiments - Experimenters Manual," Working Paper No. 21, Institute for Empirical Research in Economics, University of Zurich.

Gibbons, R. (1987): "Piece Rate Incentive Schemes," *Journal of Labor Economics*, 5(4), 413–429.

Gneezy, U., M. Niederle, and A. Rustichini (2003): "Performance in Competitive Environments: Gender Differences," *Quarterly Journal of Economics*, 118(3), 1049–1074.

Gneezy, U., and A. Rustichini (2004): "Gender and Competition at a Young Age," *American Economic Review Papers and Proceedings*, 94(2), 377–381.

Goldin, C. (1986): "Monitoring Costs and Occupational Segregation by Sex: A Historical Analysis," *Journal of Labor Economics*, 4(1), 1–27.

Grund, C., and D. Sliwka (2005): "Envy and Compassion in Tournaments," *Journal of Economics & Management Strategy*, 14(1), 187–207.

Harbring, C., and B. Irlenbusch (2003): "An Experimental Study on Tournament Design," *Labour Economics*, 10(4), 443–464.

Holt, C. A., and S. K. Laury (2002): "Risk Aversion and Incentive Effects," *American Economic Review*, 92(5), 1644–1655.

Kanemoto, Y., and W. B. MacLeod (1992): "The Ratchet Effect and the Market for Secondhand Workers," *Journal of Labor Economics*, 10(1), 85–98.

Knoeber, C. R., and W. N. Thurman (1994): "Testing the Theory of Tournaments: An Empirical Analysis of Broiler Production," *Journal of Labor Economics*, 12(2), 155–179.

Laffont, J.-J., and J. Tirole (1988): "The Dynamics of Incentive Contracts," *Econometrica*, 56(5), 1153–1175.

Lazear, E. P. (1986): "Salaries and Piece Rates," *Journal of Business*, 59(3), 405–431.

——— (2000): "Performance Pay and Productivity," *American Economic Review*, 90(5), 1346–1362.

Lazear, E. P., U. Malmendier, and R. A. Weber (2005): "Sorting in Experiments," unpublished manuscript.

Lazear, E. P., and S. Rosen (1981): "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 89(5), 841–864.

Lehrl, S., J. Merz, G. Burkard, and B. Fischer (1991): *Mehrfach-Wortschatz-Intelligenztest: MWTA-A*. perimed Fachbuch-Verlagsgesellschaft mbH, Erlangen.

MacLeod, W., and J. Malcomson (1989): "Implicit Contracts, Incentive Compatability, and Involuntary Unemployment," *Econometrica*, 57(2), 447–480.

——— (1998): "Motivation and Markets," *American Economic Review*, 88(3), 388–411.

MacLeod, W. B., and D. Parent (1999): "Job Characteristics and the Form of Compensation," in *Research in Labor Economics*, ed. by S. Polachek, vol. 18, pp. 177–242, Greenwich, Conn. JAI Press.

Nagin, D. S., J. B. Rebitzer, S. Sanders, and L. J. Taylor (2002): "Monitoring, Motivation, and Management: The Determinants of Opportunistic Behavior in a Field Experiment," *American Economic Review*, 92(4), 850–873.

Nalbantian, H. R., and A. Schotter (1997): "Productivity Under Group Incentives: An Experimental Study," *American Economic Review*, 87(3), 314–340.

Niederle, M., and L. Vesterlund (2005): "Do Women Shy Away from Competition?," NBER Working Paper No. 11474.

Oberholzer-Gee, F., and R. Eichenberger (2004): "Fairness in Extended Dictator Game Experiments," Working Paper.

Parent, D. (1999): "Methods of Pay and Earnings: A Longitudinal Analysis," *ndustrial and Labor Relations Review*, 53(1), 71–86.

Pencavel, J. H. (1977): "Work Effort, on-the-Job Screening, and Alternative Methods of Renumeration," in *Research in Labor Economics*, ed. by R. Ehrenberg, pp. 225–258, Greenwich, Conn. JAI Press.

Prendergast, C. (1999): "The Provision of Incentives in Firms," *Journal of Economic Literature*, 37(1), 7–63.

Regenstein, M., J. Meyer, and J. D. Hicks (1998): "Job Prospects for Welfare Recipients: Employers Speak Out," *Urban Institute Brief*, Series A(No. A-25).

Roth, G. (2001): *Fühlen, Denken, Handeln: Wie das Gehirn unser Verhalten steuert*. Suhrkamp Taschenbuch Wissenschaft, Frankfurt am Main.

Salop, J., and S. Salop (1976): "Self-Selection and Turnover in the Labor Market," *Quarterly Journal of Economics*, 90(4), 619–627.

Schneewind, K. A., G. Schröder, and R. B. Cattell (1983): *Der 16-Persönlichkeits-Faktoren-Test (16 PF)*. Verlag Hans Huber, Bern.

Schotter, A., and K. Weigelt (1992): "Asymmetric Tournaments, Equal Opportunity Laws and Affirmative Action: Some Experimental Results," *Quarterly Journal of Economics*, 107(2), 511–539.

Stiglitz, J. E. (1975): "Incentives, Risk, and Information: Notes Towards a Theory of Hierarchy," *The Bell Journal of Economics*, 6(2), 552–579.

Sunde, U. (2003): "Potential, Prizes and Performance: Testing Tournament Theory with Professional Tennis Data," IZA Discussion Paper No. 947.

van Dijk, F., J. Sonnemans, and F. van Winden (2001): "Incentive Systems in a Real Effort Experiment," *European Economic Review*, 45, 187–214.

Table 1: Productivity Differences

	Piece rate treatment			Tournament treatment			Revenue-sharing treatment		
	Piece rate (Median)	Fixed pay (Median)	Median test (p-value)	Tournament (Median)	Fixed pay (Median)	Median test (p-value)	Revenue sharing (Median)	Fixed pay (Median)	Median test (p-value)
Productivity:									
Indicator 1 (time needed in seconds)	8.00	28.50	0.003	11.00	26.00	0.002	13.00	24.00	0.173
Indicator 2 (time needed in seconds)	7.00	20.50	0.000	6.00	18.00	0.000	9.00	22.00	0.002
Indicator 3 (correct answers)	26.00	13.50	0.000	25.00	12.00	0.000	24.00	16.00	0.001

Notes: The table shows the median of each productivity indicator for both subgroups in all three treatments. Productivity Indicators 1 and 2 are elicited in steps 1 and 2 respectively and measure the amount of time that a person needed to solve a question with degree of difficulty 4. No monetary incentives were offered in step 1, while subjects were paid for speed in step 2 (see text for details on payment mode). Productivity Indicator 2 is censored for subjects who failed to answer the question within 30 seconds. Indicator 3 is elicited in step 3 of the experiment and measures the number of correct answers that subjects produced during a 5-minute work period in a piece-rate scheme.

Table 2: Productivity Sorting

	Piece rate treatment			Tournament treatment			Revenue-sharing treatment		
Dependent variable	1 if piece rate chosen			1 if tournament chosen			1 if revenue sharing chosen		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Productivity Indicator 1	-0.019***			-0.011**			-0.001		
	[0.006]			[0.004]			[0.002]		
	(-0.007)***			(-0.004)**			(-0.001)		
Productivity Indicator 2		-0.083***			-0.062***			-0.112***	
		[0.024]			[0.022]			[0.033]	
		(-0.031)***			(-0.025)***			(-0.043)***	
1 if Prod. Ind. 2 censored		-1.625***			-1.794***			-1.719***	
		[0.448]			[0.449]			[0.488]	
		(-0.575)***			(-0.564)***			(-0.605)***	
Productivity Indicator 3			0.107***			0.064***			0.057***
			[0.022]			[0.015]			[0.016]
			(0.039)***			(0.025)***			(0.022)***
Constant	0.712***	1.306***	-1.908***	0.276	0.826***	-1.316***	0.276*	1.605***	-0.884***
	[0.194]	[0.314]	[0.465]	[0.188]	[0.279]	[0.332]	[0.166]	[0.407]	[0.328]
Pseudo R-squared	0.148	0.162	0.323	0.058	0.167	0.192	0.004	0.157	0.149
Number of observations	80	80	80	80	80	80	80	80	80

Notes: The table shows Probit estimates of the propensity to sort into the variable payment scheme in the three different treatments. Standard errors are reported in brackets and the implied marginal effects, evaluated at the mean of observable characteristics, are shown in parentheses below the coefficient estimates. One asterisk denotes significance at the 10% level, two or three asterisks denote statistical significance at the 5% and at the 1% level respectively.

Table 3: Relative Self-Assessment and Sorting

Dependent variable	Piece rate treatment 1 if piece rate		Tournament treatment 1 if tournament		Revenue-sharing treatment 1 if revenue sharing	
	(1)	(2)	(3)	(4)	(5)	(6)
Relative self-assessment	-0.150***	-0.03	-0.142***	-0.081*	-0.099***	-0.044
	[0.036]	[0.051]	[0.034]	[0.046]	[0.032]	[0.039]
	(-0.057)***	(-0.011)	(-0.057)***	(-0.032)*	(-0.038)***	(-0.017)
Productivity Indicator 3		0.096***		0.039**		0.047**
		[0.029]		[0.020]		[0.018]
		(0.035)***		(0.016)**		(0.018)**
Constant	1.616***	-1.413	1.170***	-0.131	1.229***	-0.232
	[0.363]	[0.945]	[0.320]	[0.729]	[0.359]	[0.661]
Pseudo R-squared	0.194	0.326	0.183	0.22	0.096	0.161
Number of observations	80	80	80	80	80	80

Notes: Probit estimates. Standard errors are reported in brackets below the coefficients. Marginal effects in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%. The variable "relative self-assessment" takes values from 0 to 19 and measures a subject's estimate of the number of persons that were more productive in step 3 of the experiment. The smaller the value of the self-assessment variable is, the more productive a subject thinks he is relative to others.

Table 4: Overconfidence and Sorting

Dependent variable	Piece rate	Tournament	Revenue sharing
	(1)	(2)	(3)
Relative overassessment	0.042	0.101**	-0.002
	[0.045]	[0.050]	[0.041]
	(0.016)	(0.040**)	(-0.001)
Productivity Indicator 3	0.112***	0.133***	0.062**
	[0.031]	[0.033]	[0.031]
	(0.042***)	(0.053***)	(0.024**)
Constant	-1.970***	-2.733***	-0.977
	[0.648]	[0.697]	[0.642]
Pseudo R-squared	0.160	0.203	0.068
Number of observations	74	74	68

Notes: The table shows Probit estimates of models, in which observations are weighted by the probability that over- and under-assessment is possible (see text). Standard errors are reported in brackets below the coefficients. Marginal effects in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 5: Risk Preferences and Sorting

Dependent variable	Piece rate	Tournament	Revenue sharing
	(1)	(2)	(3)
Risk attitude	0.160*	0.330***	0.038
	[0.088]	[0.092]	[0.084]
	(0.058)*	(0.132)***	(0.014)
Productivity Indicator 3	0.105***	0.074***	0.058***
	[0.023]	[0.016]	[0.016]
	(0.038)***	(0.030)***	(0.022)***
Constant	-2.707***	-3.282***	-1.105*
	[0.663]	[0.678]	[0.595]
Pseudo R-squared	0.354	0.330	0.151
Number of observations	80	80	80

Notes: Probit estimates. Standard errors are reported in brackets and marginal effects in parentheses below coefficients; * significant at 10%; ** significant at 5%; *** significant at 1%. Coefficient estimates for models in columns labelled "trimmed" are based on observations of subjects who produced more than 9 but less than 31 answers in the 5-minute work period

Table 6: Social Preferences and Sorting

	Reciprocity			Trust		
Dependent variable	Piece rate	Tournament	Revenue sharing	Piece rate	Tournament	Revenue sharing
	(1)	(2)	(3)	(4)	(5)	(6)
1 if selfish	-0.534	-0.602	-0.999			
	[0.622]	[0.571]	[0.608]			
	(-0.206)	(-0.229)	(-0.380)			
1 if reciprocal	-0.100	-0.920**	-0.133			
	[0.458]	[0.387]	[0.352]			
	(-0.036)	(-0.353)**	(-0.051)			
Amount sent				0.007	0.002	0.003
				[0.005]	[0.004]	[0.004]
				(0.003)	(0.001)	(0.001)
Productivity Indicator 3	0.110***	0.071***	0.060***	0.110***	0.064***	0.057***
	[0.024]	[0.016]	[0.016]	[0.024]	[0.015]	[0.016]
	(0.040)***	(0.028)***	(0.023)***	(0.040)***	(0.026)***	(0.022)***
Constant	-1.825***	-0.810**	-0.763*	-2.399***	-1.466***	-1.031**
	[0.532]	[0.396]	[0.395]	[0.603]	[0.420]	[0.409]
Pseudo R-squared	0.331	0.245	0.176	0.345	0.195	0.152
Number of observations	80	80	80	80	80	80

Notes: Probit estimates. Standard errors are reported in brackets and marginal effects in parentheses below coefficients; * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 7: Proportions of Men and Women Sorting Into Variable Pay Schemes

	Piece rate	Tournament	Revenue sharing	All variable
female	45	40	46	44
male	74	58	72	68

Notes: The table shows the percentages and absolute numbers (in parentheses) of men and women who select into the variable pay schemes.

Table 8: Gender and Sorting

Dependent variable:	1 if variable pay chosen		
	(1)	(2)	(3)
1 if female	-0.626***	-0.382**	-0.181
	[0.166]	[0.183]	[0.196]
	(-0.243)***	(-0.148)**	(-0.071)
Productivity Indicator 3		0.067***	0.072***
		[0.010]	[0.010]
		(0.026)***	(0.028)***
Risk attitude			0.166***
			[0.051]
			(0.065)***
Constant	0.470***	-0.994***	-2.072***
	[0.120]	[0.241]	[0.420]
Pseudo-R-squared	0.044	0.221	0.254
Number of observations	240	240	240

Notes: Probit estimates. Marginal effects reported. Standard errors in brackets; ** significant at 5%, *** significant at 1%.

Table 9: Gender, Personality and Sorting

Dependent variable	Piece-rate treatment 1 if piece rate		Tournament treatment 1 if tournament		Revenue-sharing treatment 1 if revenue sharing	
	Men	Women	Men	Women	Men	Women
	(1)	(2)	(3)	(4)	(5)	(6)
Conscientiousness	-0.037	0.11	-0.198**	-0.114	-0.052	0.088
(Norm-orientation)	[0.075]	[0.092]	[0.088]	[0.086]	[0.066]	[0.120]
Emotional stability	0.134*	-0.029	0.053	0.125**	-0.099	0.094
	[0.069]	[0.071]	[0.074]	[0.056]	[0.061]	[0.073]
Independence	0.111	0.069	-0.025	-0.036	0.097	-0.139*
(Dominance)	[0.086]	[0.062]	[0.081]	[0.075]	[0.065]	[0.078]
Readiness to take decisions	-0.082	-0.018	-0.005	-0.011	0.084	0.031
	[0.086]	[0.064]	[0.087]	[0.071]	[0.065]	[0.057]
Extroversion	-0.054	-0.144*	-0.129	0.002	-0.007	0.047
	[0.066]	[0.085]	[0.088]	[0.065]	[0.052]	[0.064]
Pseudo R-squared	0.240	0.087	0.168	0.123	0.119	0.113
Number of observations	40	35	36	38	36	37

Notes: Probit Estimates. Marginal effects reported. Standard errors in brackets; ** significant at 5%, *** significant at 1%. Observations of subjects who indicated that their answers are not reliable are discarded.

Table 10: Effort, Stress, and Exhaustion

	Before sorting decision			After sorting decision		
	Piece rate	Fixed	M-W test	Piece rate	Fixed	M-W test
	(Mean)	(Mean)	(p-value)	(Mean)	(Mean)	(p-value)
	(1)	(2)	(3)	(4)	(5)	(6)
Effort	5.63	5.50	0.559	6.00	4.25	0.000
Stress	5.44	5.53	0.757	5.60	3.56	0.000
Exhaustion	2.96	2.59	0.448	4.00	2.59	0.001
Number of observations	48	32		48	32	

(a) Effort, Stress and Exhaustion in Piece-Rate Treatment

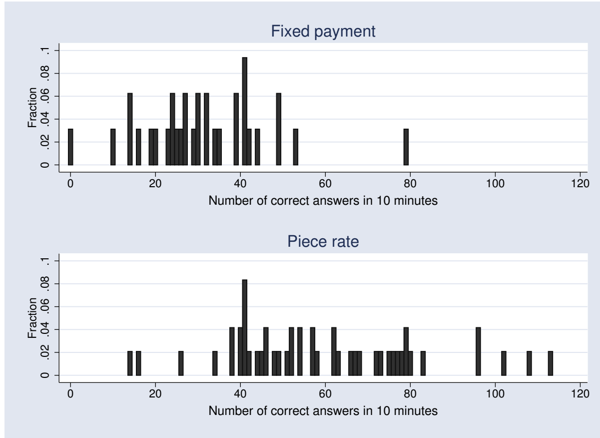
	Before sorting decision			After sorting decision		
	Tournament	Fixed	M-W test	Tournament	Fixed	M-W test
	(Mean)	(Mean)	(p-value)	(Mean)	(Mean)	(p-value)
	(1)	(2)	(3)	(4)	(5)	(6)
Effort	5.54	5.39	0.442	6.15	4.76	0.000
Stress	5.54	5.41	0.854	5.74	3.98	0.000
Exhaustion	2.90	2.85	0.749	3.36	3.29	0.773
Number of observations	39	41		39	41	

(b) Effort, Stress and Exhaustion in Tournament Treatment

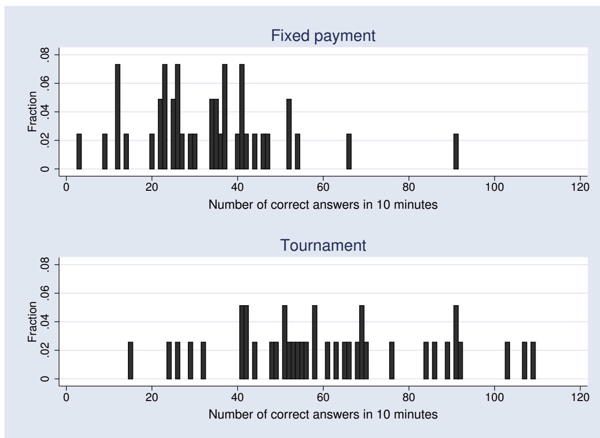
	Before sorting decision			After sorting decision		
	Revenue sharing	Fixed	M-W test	Revenue sharing	Fixed	M-W test
	(Mean)	(Mean)	(p-value)	(Mean)	(Mean)	(p-value)
	(1)	(2)	(3)	(4)	(5)	(6)
Effort	5.43	5.30	0.703	5.43	4.45	0.001
Stress	5.36	5.45	0.500	5.40	3.79	0.000
Exhaustion	2.43	2.18	0.518	3.60	2.52	0.006
Number of observations	47	33		47	33	

(c) Effort, Stress and Exhaustion in Revenue-Sharing Treatment

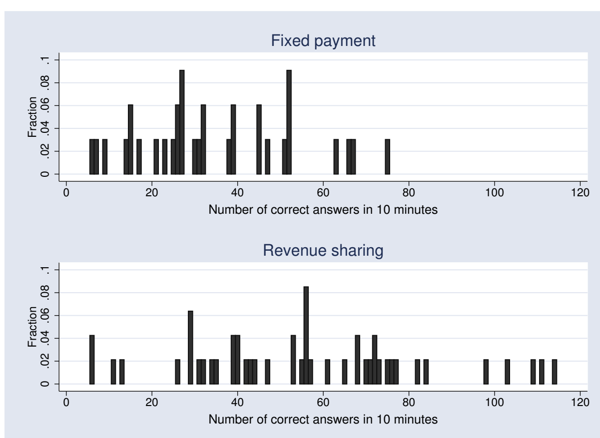
Figure 1: Output of Self-Selected Subjects in Different Compensation Schemes



(a) Piece-rate treatment



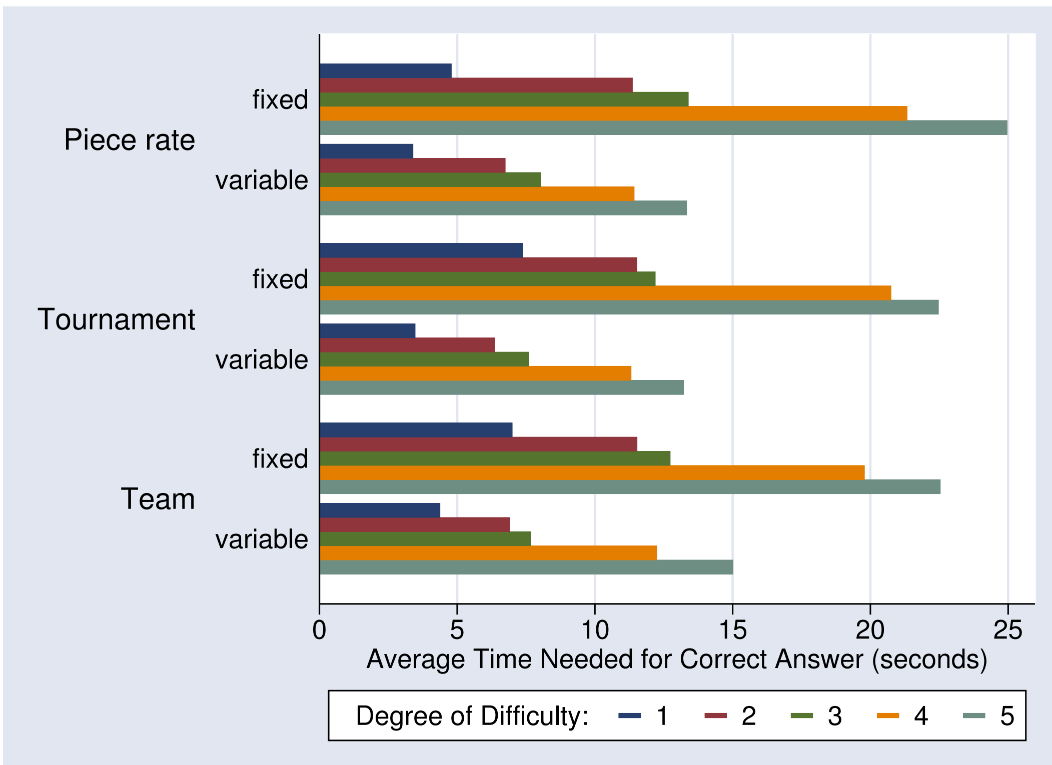
(b) Tournament treatment



(c) Revenue-sharing treatment

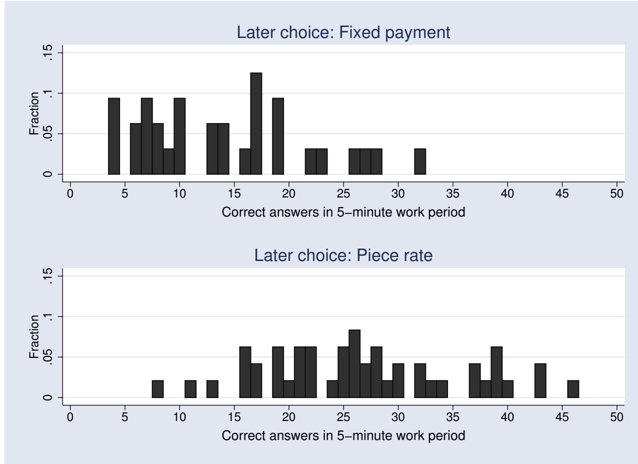
Notes: Each panel of the figure plots, for a particular treatment, two histograms of individual output (measured as the number of correct answers during the total working time of ten minutes), one for each of the self-selected groups of subjects: The upper histogram shows the output distribution of workers who selected into the fixed-payment contract, and the lower histogram of a panel shows the output distribution of workers who selected into the variable payment contract. Panel (a) shows output histograms for the piece-rate treatment, Panel (b) those that arose in the tournament treatment, and Panel (c) plots output histograms from the revenue-sharing treatment.

Figure 2: Performance and Task Difficulty

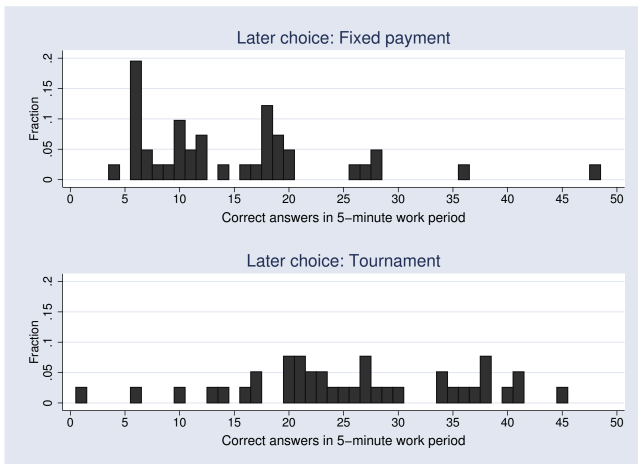


Notes: The figure shows, for each treatment, how much time (in seconds) subjects working in a particular self-selected regime need on average to solve a question of a given degree of difficulty.

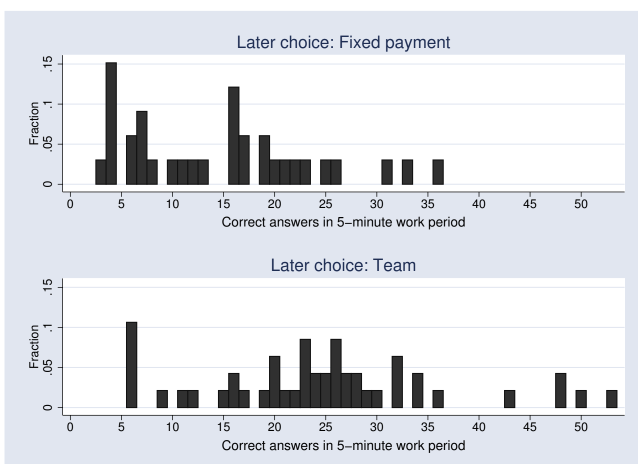
Figure 3: Productivity of Subjects before Self-Selection Into Incentive Contract



(a) Piece-rate treatment



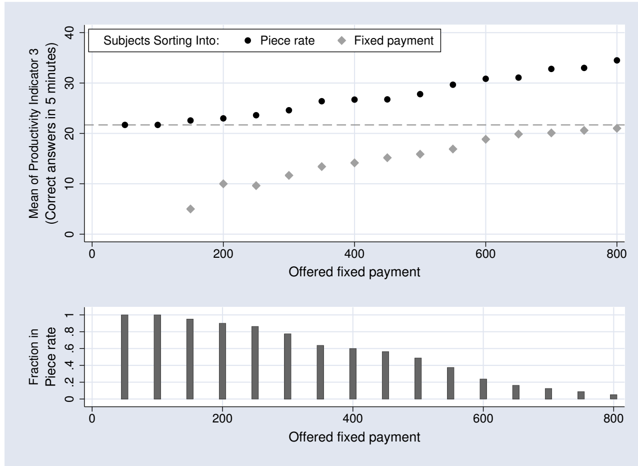
(b) Tournament treatment



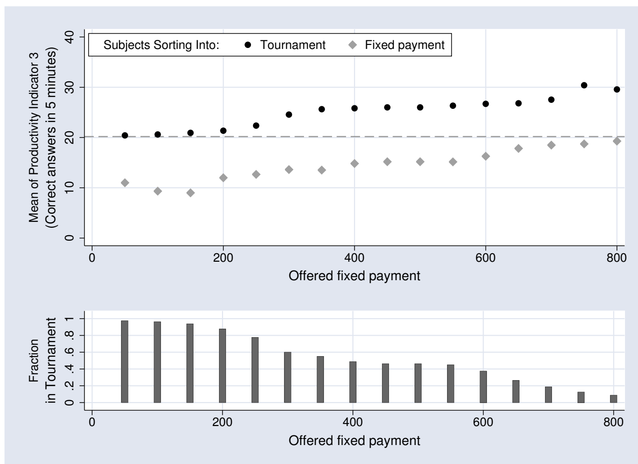
(c) Revenue-sharing treatment

Notes: Each panel of the figure plots histograms of Productivity Indicator 3, which was elicited in step 3 of the experiment and which measures the number of correct answers during a 5-minute work period. The upper histogram always shows the distribution of individual productivity for subjects who subsequently chose the fixed payment alternative, while the lower histogram of a panel always shows the productivity distribution among subjects who subsequently preferred the variable payment alternative. Panel (a) refers to the piece-rate treatment, and Panel (b) and Panel (c) to the tournament treatment and revenue-sharing treatment respectively.

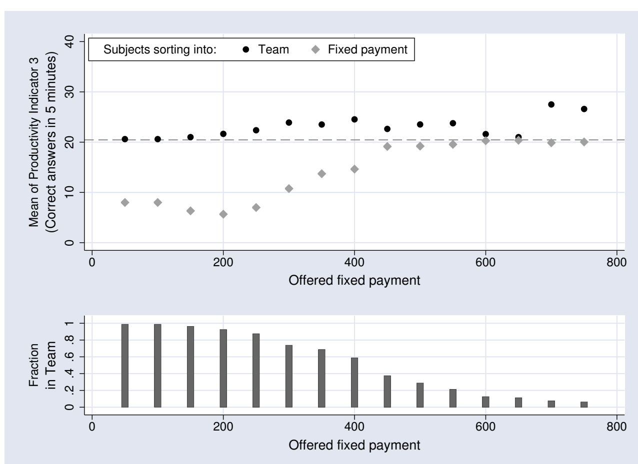
Figure 4: Fraction of Subjects Opting for Variable Pay and Average Productivity of Sorted Subjects



(a) Piece-rate treatment



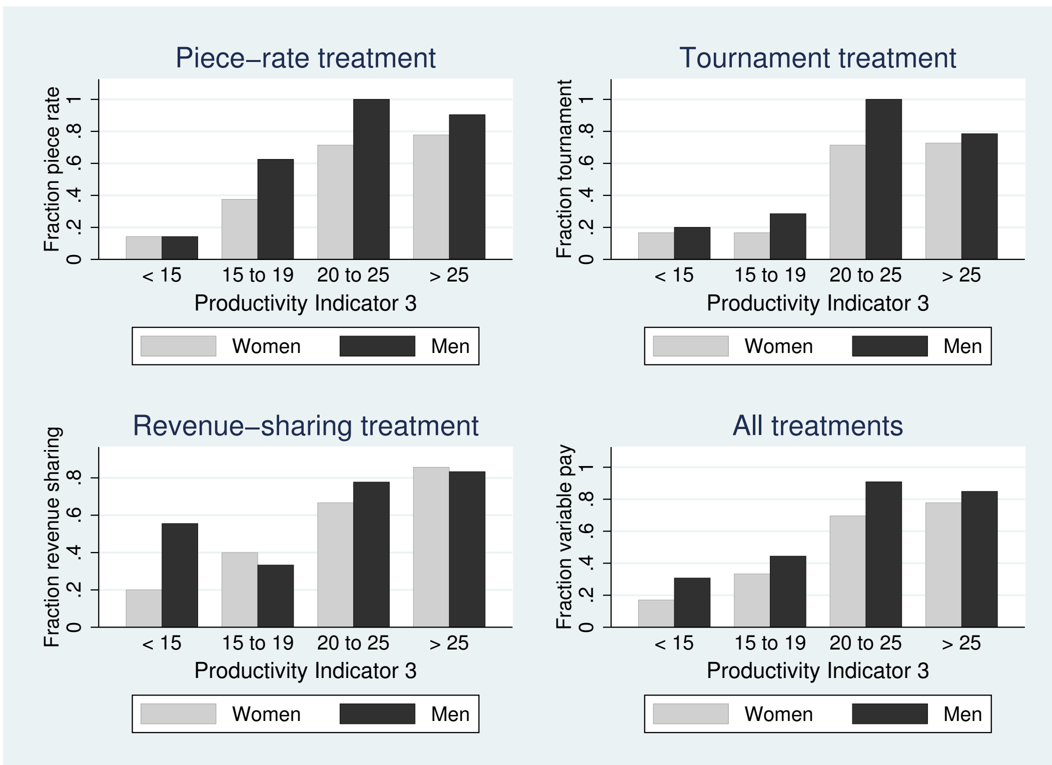
(b) Tournament treatment



(c) Revenue-sharing treatment

Notes: The upper graph of each panel shows average productivity, measure by Productivity Indicator 3, among subjects who would sort into the respective variable payment scheme at a particular fixed payment alternative. The lower graph of each panel displays the number of subjects who would opt for variable compensation at a given offered fixed payment alternative. Panel (a) refers to the piece-rate treatment, Panel (b) to the tournament treatment and Panel (c) to the revenue-sharing treatment.

Figure 5: Gender and Sorting



Notes: The figure shows what fraction of men and women with a particular productivity level selects into the variable payment scheme.

Figure A.1: Design of the Experiment

Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8	Step 9	Step 10	Step 11	Step 12
Productivity Indicator 1	Productivity Indicator 2	Productivity Indicator 3	Effort questions	Relative self-assessment	Sorting decision	Sorting with different fixed payment alternatives	Working time	Effort questions	Social preferences	Risk preferences	Question-naires
Calculate one problem of difficulty level 4 as fast as possible	Calculate one problem of difficulty level 4 as fast as possible (paid)	Piece rate with 10 P per correct answer 5 minutes	How much effort have you exerted? How stressed did you feel? How exhausted did you get?	How many people (out of 20) solved more question better than you did? Paid correct: 100 P +/-1: 50 P	a) **Piece rate:** 10 P per correct answer b) **Tournament:** Winner is who has more correct answers Winner gets 1300P Loser gets 0P c) **Revenue sharing:** (Sum of output) *10 divided by 2 vs. **Fixed payment:** 400 P independent of output	Varying the fixed payment alternatives	10 minutes Piece rate, Tournament, Revenue sharing or Fixed payment	How much effort have you exerted? How stressed did you feel? How exhausted did you get?	2-player, sequential trust game Endowment of 120 Transfer of 1st mover tripled Contingent response method for 2nd mover Role reversal No information	Choice between L(400, 0; 0.5) and 15 safe options 25, 50, …, 375 One alternative randomly chosen	Risk-preference questions Attitudes"Big Five" Socioeconomics Math and high-school grades