



GOVERNANCE AND THE EFFICIENCY
OF ECONOMIC SYSTEMS
GESY

Discussion Paper No. 360

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

Thomas Dohmen*
Armin Falk**

*ROA, Maastricht University
**University of Bonn

May 2011

Financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15 is gratefully acknowledged.

Sonderforschungsbereich/Transregio 15 · www.sfbtr15.de
Universität Mannheim · Freie Universität Berlin · Humboldt-Universität zu Berlin · Ludwig-Maximilians-Universität München
Rheinische Friedrich-Wilhelms-Universität Bonn · Zentrum für Europäische Wirtschaftsforschung Mannheim

Speaker: Prof. Dr. Urs Schweizer · Department of Economics · University of Bonn · D-53113 Bonn,
Phone: +49(0228)739220 · Fax: +49(0228)739221

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

By THOMAS DOHMEN AND ARMIN FALK*

This paper studies the impact of incentives on worker self-selection in a controlled laboratory experiment. Subjects face the choice between a fixed and a variable payment scheme. Depending on the treatment, the variable payment is a piece rate, a tournament or a revenue-sharing scheme. We find that output is higher in the variable pay schemes (piece rate, tournament, and revenue sharing) compared to the fixed payment scheme. This difference is largely driven by productivity sorting. In addition personal attitudes such as willingness to take risks and relative self-assessment as well as gender affect the sorting decision in a systematic way. Moreover, self-reported effort is significantly higher in all variable pay conditions than in the fixed wage condition. Our lab findings are supported by an additional analysis using data from a large and representative sample. 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 individual characteristics.

JEL: J3, M52, C91, D81, J16

Keywords: Sorting, Incentives, Piece Rates, Tournament, Revenue-Sharing, Risk Preferences, Social Preferences, Gender, Experiment, Field Evidence

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 underestimates the importance of worker self-selection, i.e., the possibility that agents with different individual characteristics feel attracted by different pay schemes and therefore systematically self-select into particular firms and organizations. In the presence of self-selection performance is likely to depend not only on the the incentive effect *per se*, but also on the sorting effect, which affects the composition of the workforce. A few studies (e.g., Edward P. Lazear, 2000?) indicate that productivity sorting contributes to output differences between different incentive systems. Little is known empirically, however, about the nature of this selection process along other dimensions that are crucial to an organization's success such as workers' preferences and attitudes.

Field data often lack important information on workers' preferences and motives, and confounding factors impede causal inference. This paper therefore explores the driving forces of self-selection in a controlled laboratory environment. We address the following questions: Which personal characteristics beyond individual productivity differences provoke workers to self-select into variable instead of fixed pay contracts? In particular, how do relevant characteristics like risk aversion, relative self-assessment, 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

* Dohmen: ROA, Maastricht University; Address: Maastricht University, Research Centre for Education and the Labour Market (ROA), P.O. Box 616, 6200 MD Maastricht, The Netherlands; Email: t.dohmen@roa.unimaas.nl. Falk: University of Bonn; Address: University of Bonn, Department of Economics, Adenauerallee 24-42, D-53113 Bonn, Germany. Email: armin.falk@uni-bonn.de. Acknowledgements

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. We then elicit further individual characteristics that may be relevant for the sorting decision. Finally, we obtain self-reported measures of work effort, stress and exhaustion. The work task consists of multiplying one-digit numbers by 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 on the one hand and either a piece rate, a tournament, or a revenue-sharing scheme on the other hand. These three forms of variable pay constitute the most important 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 results reveal the importance of multi-dimensional sorting. We first establish that output in all variable payment schemes is higher than output under the fixed wage regime. This output difference is mainly attributable to productivity sorting, which is strong and present 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. Our results further show that relative self-assessment plays an important role for sorting into tournaments, which makes sense as payments in tournament schemes depend on relative performance. Another important driver of self-selection is a subject's attitude towards risk: the likelihood that subjects prefer the piece rate or the tournament is higher the less risk averse they are. This finding reflects the fact that the fixed payment yields a safe payoff whereas earnings variation renders the variable pay alternative risky. We also show that women are less likely to select into variable pay schemes than men. This is mostly explained by differences in risk attitudes and productivity between men and women. In an additional analysis we show that the extent to which personal characteristics affect the sorting decision depends on whether a subject is a "marginal" type, i.e., someone whose decision is on the fence. We either use the location in the productivity distribution or response times in order to characterize marginal types. It is plausible to assume that those who are on the fence take a longer time to make their sorting decision. It turns out that sorting patterns do in fact vary by this criterion. For example, risk attitudes play a bigger role in the piece-rate treatment for subjects who are marginal, i.e., who either contemplate longer or are close to the productivity threshold that makes them indifferent between fixed wages and piece rates. Differential patterns for marginal and non-marginal types are also observed in the tournament and revenue-sharing treatments. For example, in the tournament treatment risk attitudes matter for marginal types but not for non-marginal types who base their decision on productivity and relative self-assessment.

On top of the observed sorting patterns we show that self-reported effort and work output vary significantly with different 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.

In Section III we compare our sorting results with sorting that takes place in the German labor market. This complementary analysis is based on data from the German Socio-Economic Panel Study (SOEP), a large panel survey that is representative of the resident adult population. We estimate how productivity, risk attitudes, social preferences and gender affect the probability of working under variable pay. The picture that emerges

is very similar to the laboratory results: More productive and risk tolerant workers are more likely to work for performance pay while women are less likely to work under variable payment schemes. This result in combination with our controlled lab evidence provides a powerful confirmation of the importance of multi-dimensional sorting.

The literature on optimal incentives has shown that characteristics of the production processes and the information structure affect optimal employment contracts.¹ Our results indicate that organizations should, in addition, take into account the interaction of incentives and multi-dimensional sorting when deciding on the design of the incentive system. This follows from the fact that worker characteristics and preferences affect the success of firms. This is quite obvious for productivity. But also risk or social preferences, as well as self-assessments may have an important influence on a firm's success. For example, if an investment company attracts relatively risk loving and overconfident fund managers, this will most likely affect the company's portfolio strategy. As another example, social preferences can be relevant for reducing free-riding in teams and may therefore positively affect output (see, e.g., Ernst Fehr and Simon Gächter, 2000?). Of course, many of the discussed worker attributes are typically unobservable during the hiring process. In this sense our results suggest that firms may use incentive schemes as screening devices to attract particular types of workers (Joanne Salop and Steven Salop, 1976?). 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 imply that changing the pay system sets off sorting processes beyond productivity sorting. This change in the workforce composition might affect several procedures inside the firm and change the entire work environment or firm culture (see, e.g., Michael Kosfeld and Ferdinand von Siemens, 2007?). It is also important to realize that introducing variable pay in certain jobs that are predominantly characterized by fixed wage schemes, such as the public sector, are likely to reduce job satisfaction among incumbent workers who prefer the previous wage system, as revealed by their past choice.

Showing the relevance of sorting also underlines the methodological difficulties associated with testing contract theory with field data (Canice Prendergast, 1999?; Pierre-André Chiappori and Bernard Salanié?, 2003). When 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 made in the theoretical analysis by Lazear (1986)? and empirically shown in a well-known case study of a firm that changed from fixed wages to piece rates (Lazear, 2000)?. Ignoring the sorting effect would imply a dramatic overestimation of the incentive effect. Our results confirm Lazear's (2000) finding that productivity sorting is an important source of output differences between piece-rate and fixed-wage pay schemes, but also highlight that productivity sorting is a key driver of positive output effects of other variable pay schemes, such as tournaments and revenue sharing. Moreover, our results point to another potential confounding factor in testing contract theory: Pref-

¹Early work (e.g., Joseph E. 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., Charles Brown, 1990, 1992, 1994??; Claudia Goldin, 1986?; Daniel Parent, 1999?; John H. Pencavel, 1977?). Lazear and Sherwin 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., Jean-Jacques Laffont and Jean Tirole, 1988?), which also highlights the role of future commitment (see David P. Baron and David Besanko, 1984?; Robert Gibbons, 1987?; Yoshitsugu Kanemoto and W. Bentley MacLeod, 1992?). For evidence on the interplay between job characteristics and the incidence of particular compensation contracts see also MacLeod and Parent (1999)?.

erence and self-assessment sorting. It is well known that optimal contracts depend on risk preferences. In light of our findings, the composition of preferences and individual attitudes in a given pool of agents is likely to be endogenous.

Ruling out endogeneity that stems from selection is an important 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 Our experimental design shows that experiments can also be used to study sorting in a controlled way.³ A similar approach was used in the studies by Bram C. Cadsby, Fei Song, and Francis Tapon (2007)? and Eriksson, Sabrina Teyssier, and Villeval (2009)? 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 face in the experiment by Muriel Niederle and Lise Vesterlund (2007)?. Based on the finding that women perform worse in the presence of men in competitive environments (Uri Gneezy and Aldo 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. Section II presents the results. We first discuss the output effects of different incentive schemes. Then we present evidence on the importance of sorting. Finally, we discuss the effect of incentives on the provision of effort. In Section III we provide additional survey evidence and discuss implications of our results.

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

The ideal data set for studying how individual characteristics affect the sorting decision into different incentive schemes 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, 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 relative self-assessment. Second, workers are typically exposed to a mix of explicit

²Using random treatment assignment, tournament incentives have been studied, e.g., by Clive Bull, Andrew Schotter and Keith Weigelt (1987)?, Schotter and Weigelt (1992)?, Armin Falk and Fehr (2002)? and Christine Harbring and Bernd Irlenbusch (2003)?. The lab evidence on tournaments is complemented by field studies on corporate tournaments (Michael L. Bognanno, 2001?), tournaments in agricultural production (Charles R. Knoeber and Walter N. Thurman, 1994?) and sports tournaments (e.g., Ronald G. Ehrenberg and Bognanno, 1990?; Sue Fernie and David Metcalf, 1999?; and Uwe Sunde, 2009?). The incentive effects of piece rates have been experimentally investigated, e.g., by Bull, Schotter and Weigelt (1987)? and Frans van Dijk, Joep Sonnemans and Frans van Winden (2001)?, while team incentives have been studied, e.g., by Haig Nalbantian and Schotter (1997)?. The impact of incentives has also been studied in field experiments, e.g., Oriana Bandiera, Iwan Barankay, and Imran Rasul (2005)? and Daniel S. Nagin et al. (2002)?.

³In a similar vein sorting has been studied, e.g., in a market entry game (Colin F. Camerer and Dan Lovo, 1999)?, in simple bargaining games (Felix Oberholzer-Gee and Reiner Eichenberger, 2008?; and Lazear, Ulrike Malmendier and Roberto A. Weber, 2005?), the gift-exchange game (Tor Eriksson and Marie-Claire Villeval, 2008?) or the prisoner's dilemma game (Iris Bohnet and Dorothea Kübler, 2005?).

and implicit incentives, which complicates an accurate characterization of the incentives that actually prevail in a given work environment. Suppose, for example, that the researcher observes that a firm has established piece-rate contracts. This neither precludes the possibility that workers are also motivated by the threat of dismissal or the chance of being promoted, nor that they are simultaneously confronted with additional incentives that directly affect remuneration, for example, bonus payments or team incentives like profit sharing. Likewise, workers without explicit performance pay contracts might face work incentives stemming from implicit contracts and repeated game effects (MacLeod and James M. Malcomson, 1989?, 1998?). Third, individual output measures are often not available or are fraught with measurement error. 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 (see Section III and Falk and James J. Heckman, 2009?). 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 rule out any mix of different implicit or explicit incentives. Finally, since the sorting decision takes place immediately, timing is not an issue.

A. *The Work Task*

The work task implemented in our experiment consists of multiplying one-digit numbers by two-digit numbers. This “real effort” task implies that subjects have to actually work⁴ 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 (Gerhard Roth, 2001?). Depending on the chosen numbers, the difficulty level of multiplying one-digit numbers by 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.⁵ As we will see below, solving more difficult problems is more time-consuming.

All problems were presented to subjects on computer screens (see Appendix). They had to 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

⁴This 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., René Fahr and Bernd 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, Niederle, and Rustichini (2003)? who had subjects solve mazes at the computer and Falk and Andrea Ichino (2006)? who asked subjects to stuff letters into envelopes.

⁵Examples 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: $7 \cdot 89$.

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 or she had answered correctly.

B. 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 different steps, but they did not know what these steps would look like. 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 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 (Productivity Indicator 2) 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 150 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.; subjects who did not come up with the correct answer in 30 seconds earned nothing. A clock on the screen informed about how many of the 30 seconds had elapsed.

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. Since our third productivity indicator is based on a sequence of problems of different difficulty levels (i.e., just like in the actual work task in step 8 of the experiment) rather than on a single question, it is most informative for the work environment under study. We will therefore focus on Productivity Indicator 3 in Section II.

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 you exert? How stressed did you feel? How exhausted did you get? Answers to these questions were given on a seven point Likert 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 relative overassessment. The question subjects had to answer reads as follows: How many of the other 19 participants solved more questions 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 zero points otherwise. Subjects were not informed about their true rank in the distribution before 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 , guarantees the payment of 400 points independent of output x , the number of correctly answered problems, i.e.,

$$(1) \quad w_i^F = 400.$$

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 and revenue-sharing contracts. In the piece-rate treatment, the alternative contract paid a piece rate 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

$$(2) \quad w_i^{PR} = 10 \cdot x_i.$$

In the tournament treatment, 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 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 w^T for player i is given by

$$(3) \quad w_i^T = \begin{cases} 1300 & \text{if } x_i > x_j, \quad i \neq j, \\ 1300 \text{ with probability } 0.5 \text{ and } 0 \text{ with probability } 0.5 & \text{if } x_i = x_j, \\ 0 & \text{otherwise} \end{cases}$$

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 revenue 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 ten points for each correctly answered problem. A team's revenue was then divided equally among the two team members. The compensation for player i in the

revenue-sharing condition w^{RS} is hence given by

$$(4) \quad w_i^{RS} = 10 \cdot \frac{x_i + x_j}{2}.$$

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 $\{50, 100, 150, \dots, 800\}$ points. These hypothetical choices reveal information about sorting patterns at more or 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 Joyce Berg, John Dickhaut, and Kevin McCabe, 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 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 Charles A. Holt and Susan K. 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.

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 Socio-Economic Panel Study (see also Section III).⁶ Dohmen et

⁶The exact wording of the question (translated from German) is as follows: How do you see yourself:

al. (2009)? 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.

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 developed by Siegfried Lehrl et al. (1991)?, and a personal attitudes test developed by Hermann Brandstätter (see Brandstätter, 1988?).⁷

C. Procedural Details

The experiment was computerized using the software z-Tree (Urs 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 eighteen sessions, six sessions in each of the three treatments. A total of 360 subjects participated. We invited the same number of females and males in each session and ended up with 181 female and 179 male participants.⁸ 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 1.30 US Dollar at that time). Average earnings were 21.20 Euro.

II. Results

In this section we present the main results. In Section A, we start by investigating whether subjects who opt for a variable pay contract produce more than subjects who prefer to work for a fixed payment. In Section B we focus on the role of sorting.

A. 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

“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’.”

⁷This so called 16 PA test is a short form of the German-language version of Raymond B. Cattell's sixteen personality factor questionnaire (16 PF), an internationally well established personality assessment, that produces five dimensions of personality (the so-called “Big Five”). The German-language version of the 16 PF was developed by Klaus A. Schneewind, Gundo Schröder and Raymond B. Cattell (1983)? and contains 192 items that compass sixteen primary scales of personality. These primary scales produce five independent secondary factors (the so-called “Big Five”), which are commonly labeled as conscientiousness, neuroticism, openness to experience, agreeableness and extroversion. Brandstätter's short personal attitudes test presents subjects with 32 conflictive adjective pairs, which describe traits. For each adjective pair, subjects indicate how they would assess themselves on a 9-point scale that is spanned by the conflictive adjectives. Based on these 32 ratings, 16 primary factors are constructed. We then use principal component analysis to extract the five principal components that describe personality.

⁸We invited 12 men and 12 women to each session. The first 20 subjects who showed up at the lab participated in the experiment. The other subjects received a show-up fee and were asked to leave the laboratory.

in 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. Figures 1 and 2 in fact show that output in all variable pay schemes is higher than output under the fixed wage regime. Charts (a), (b) and (c) of Figure 1 depict, for each of the three treatments, the cumulative distributions of realized output during the 10-minute work period (step 8 of the experiment) separately for subjects who have self-selected into the fixed payment scheme and those who opted for variable pay. In all three treatments the output distribution in the variable pay condition statistically dominates the output distribution in the respective fixed payment condition. Subjects with a piece-rate contract solved on average 60.59 problems (sample standard deviation: 21.81) compared to 29.51 (std.dev. 14.22) problems solved by subjects who worked for the fixed payment in the same treatment. The respective numbers in the tournament treatment are 61.08 (std.dev. 22.34) versus 36.08 (std.dev. 18.81) and in the revenue-sharing treatment 57.93 (std.dev. 28.02) versus 33.75 (std.dev. 16.93). Output differences between fixed and variable pay schemes in all treatments are statistically significant at any conventional level (Wilcoxon rank-sum test, p -value < 0.0001).⁹

Figure 2 restates this result in a different way. 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, the time needed to solve a problem increases in the level of difficulty (level 1 to 5). Moreover, 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 explained by the fact that subjects in the fixed payment condition simply need longer to calculate problems, but also that the error rate, which generally rises with the degree of difficulty, is higher for subjects in the fixed payment schemes.¹⁰

B. Sorting

In order to explain the output differences discussed above and to understand the interaction of incentive systems and sorting, we now analyze how sorting is affected by personal attributes. We start with productivity sorting.

⁹Firms may not only be concerned about level effects of output but also about differences in the variance of output. We find that differences in output variances for fixed wage workers are not statistically significant at the 5 percent significance level for any treatment comparison. Likewise, the variance of output for workers who opted for the variable pay in the piece-rate treatment does not differ significantly from the output variance among tournament participants (H_0 : variances are the same, p -value < 0.8427). However, the difference in output variances among tournament participants and participants in the revenue-sharing scheme is marginally statistically significant (p -value < 0.0714), and the variance of output among workers who opted for the variable pay in the revenue-sharing treatment is significantly higher than the output variance of variable-pay workers in the piece-rate treatment (p -value < 0.0337).

¹⁰Subjects solved 95.4 percent of problems on their first attempt when the degree of difficulty was equal to 1, 91.4 percent when the degree of difficulty was 2, 87.8 percent when it was 3, 81.0 percent when the degree of difficulty was 4 and 77.5 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.

PRODUCTIVITY

If 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 increases in the level of the fixed payment, and it decreases in the attractiveness of the piece rate.¹¹ If the difference between optimal effort costs under piece rates and fixed wages is sufficiently small, risk-neutral subjects, who expect to solve more than 40 problems in 10 minutes, optimally choose the piece rate. This is the case for subjects who produced more than 20 correct answers during the 5-minute work period and who expect that they can solve twice as many problems in 10 minutes than in 5 minutes.

In the tournament, a similar productivity sorting pattern is plausible but not as obvious as in the piece-rate treatment. The reason is the strategic nature of tournaments: A risk-neutral subject optimally participates in the tournament if the winning prize of 1300 points multiplied by the probability of winning exceeds 400 points plus the value of the disutility that results from providing higher effort in the tournament than in the fixed payment.¹² Thus the sorting decision does not only depend on own productivity 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. Productivity sorting is also more likely the less important luck is relative to differences in ability.¹³ In our experiment, “luck” could be a sequence of easy problems, or subjective productivity shocks such as mood effects or other psychological aspects of motivation or distraction. Subjects could evaluate the importance of luck prior to their sorting decision since they learned about the structure of problems in the first five minutes. They knew that the task and the difficulty level of questions would be the same in the 10-minute work period. Productivity differences

¹¹More formally, this can be shown as follows: Assume that an individual’s output, x_i , depends on his productivity, π_i , which is a function of ability, θ_i , and effort, $e_i \geq 0$, i.e., the production function is given by $x_i = \pi(\theta_i, e_i) + \varepsilon_i$, where $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$, $\pi^\theta, \pi^e > 0$ and $\pi^{e\theta} \geq 0$. Subjects’ ability is continuously distributed on the interval $[\underline{\theta}, \bar{\theta}]$ according to the cumulative distribution function $F(\theta)$. Assume that a subject’s utility depends positively on the wage w and negatively on effort e according to the utility function $u(w, e) = w - c(e)$ with $u^w > 0$, and $u^c < 0$ since $c^e > 0$. Expected utility in the fixed wage regime with $w^F = \alpha > 0$ is maximized if the minimum required level of effort denoted by e^{min} is exerted, because remuneration is independent of effort in the fixed payment regime and $c^e > 0$. 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. Expected utility under piece rates is maximized when the optimal effort level e^* , which satisfies $\frac{\delta c(e)}{\delta e} = \beta \frac{\delta \pi}{\delta e}$. Risk-neutral subjects opt for the contract that results in higher utility, so that the piece-rate contract is preferred when productivity exceeds the productivity threshold, $\hat{\pi}$, which is given by $\hat{\pi} = \frac{\alpha + c(e^*) - c(e^{min})}{\beta}$. The term $c(e^*) - c(e^{min})$ captures the disutility that results when effort is raised from e^{min} to e^* . Note that $\hat{\pi}$ increases in α and decreases in β .

¹²In the framework introduced in footnote 11, a risk-neutral subject optimally participates in the tournament if $\gamma \cdot \text{Prob}\{\pi_i(\theta_i, e_i^*) - \pi_j(\theta_j, e_j^*) > e_j - e_i\} \geq \alpha - c(e^{min}) + c(e_i^*)$.

¹³For example, if luck is absent, i.e., $\sigma_\varepsilon = 0$ in the production function, and ability is continuously distributed on a closed interval $[\underline{\theta}, \bar{\theta}]$, a more able contestant has an optimal effort response function that ensures winning the tournament against a less able competitor. Since the most able subject always wins — and consequently enters the tournament — it is not optimal for a less productive person to compete. Entering the tournament is a weakly dominant strategy for the most able subject as he receives the outside option when no tournament takes place. In this setting no tournament takes place, as only the most productive individual optimally opts for tournament incentives. 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 $\underline{\theta} < \hat{\theta} < \bar{\theta}$ sort into the tournament and less able subjects select into the fixed payment scheme, may exist for intermediate cases.

are likely to dominate luck in determining output and thus the likelihood of winning. As a consequence, we expect an outcome in which productive workers are more likely to participate in the tournament than less productive workers.

As in the tournament treatment, the prediction for productivity sorting in the revenue-sharing treatment depends on distributional assumptions. It is possible in theory to fix parameters such that either all subjects are expected to join the revenue-sharing contract, or some are expected to join, either with or without a unique threshold. Note, however, that for highly productive types, who can attain higher utility than under fixed wages even if their team partner does not produce anything, have a dominant strategy to sort into the revenue-sharing scheme. Abstracting from effort costs, the corresponding critical output is 80 correct answers during the 10-minute working time.¹⁴ We therefore expect average productivity to be higher among revenue-sharing participants than among subjects in the fixed wage scheme. Taken together, productivity sorting is likely to occur in all treatments, and especially in the piece-rate treatment.

Our results confirm systematic productivity sorting in all treatments. On average, the more productive a worker, the more likely he self-selects into the variable pay scheme. Support for this result comes from Figures 3 and 4. Figure 3 contains three charts, each of which compares the cumulative distributions of productivity (measured by Productivity Indicator 3) of subjects who sorted into the fixed payment scheme and of subjects who sorted into the variable payment scheme in a particular treatment. The fractions of subjects who self-select into the variable pay scheme are 60.83 percent in the piece-rate treatment, 50.0 percent in the tournament and 63.33 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 should therefore expect to produce more than 80 correct answers in the 10-minute work period, sort into the revenue-sharing scheme.

Wilcoxon rank-sum tests verify that the productivity differences 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 27.3 correct answers compared to an average productivity of 13.8 of subjects who sort into a fixed-payment contract. Corresponding numbers are 26.0 versus 16.4 for the tournament treatment, and 25.9 versus 14.9 for the revenue-sharing treatment. Since all three productivity measures are highly significantly correlated (Spearman rank correlations, p-values < 0.0001), it is not surprising that our productivity result holds regardless of the productivity indicator used.¹⁵

¹⁴Along the lines of footnote 11 it can be shown that subjects whose team partner does not produce any output optimally opt for the revenue-sharing contract if their own productivity exceeds $\frac{2(\alpha - c(e^{min}) + c(e^*))}{\beta}$.

¹⁵The Spearman rank correlation between Indicators 3 and 1 is -0.647 (p-value < 0.0001), i.e., individuals who are faster in entering the correct answer in step 1 also solve more problems during the 5 minutes in step 3. The Spearman rank correlation between Indicators 3 and 2 is -0.615 (p-value < 0.0001) among the 290 individuals with uncensored observations, while the correlation between indicators 1 and 2 is 0.521 (p-value < 0.0001). Productivity differences, measured by indicators 1 and 2, between those who opt for the fixed payment and those who prefer the variable payment are highly significant in all three treatments. The median subjects in the piece-rate treatment, tournament treatment and revenue-sharing treatment who opt for the fixed payment need 27, 23 and 20.5 seconds respectively in order to solve the problem in step 1 of the experiment (Productivity Indicator 1), while median subjects who opt for the variable pay need 7, 8, and 12 seconds respectively. Similarly the median time needed to solve the prob-

The fact that the productivity indicators are also highly significantly correlated with a measure of verbal IQ, the *Abitur* grade (*Abitur* is an exam that comes at the end of university-track high school in Germany and is a prerequisite for attending university; the grades range from 4.0, the worst grade, to 1.0, the best grade), and the final math grade in high school (ranging from 1, the worst grade, to 15, the best grade) suggests that variable payment schemes attract generally more able subjects for the work task in our experiment.¹⁶

An important implication of productivity sorting is that average productivity of a selected group depends on the relative attractiveness of the contract alternatives. 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 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 lowest fixed wage a subject just prefers over the variable payment is positive and highly significant (Spearman rank correlations; p-values < 0.001 in all treatments).

Figure 4 shows the sorting pattern in all treatments for different (hypothetical) fixed payments according to 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 higher the lower the fixed payment displayed in steps of 50 points on the horizontal axis. For example, when the fixed payment is 50, all workers prefer the piece rate, while 60.8 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 this treatment. Since all workers prefer the piece rate to fixed payments for very low fixed wages, 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.

lem in step 2 of the experiment (Productivity Indicator 2) is 16, 15.5 and 19.5 seconds for subjects who opt for the fixed payment and 7, 5.5 and 8 seconds for those who select into the variable pay scheme in the piece-rate, tournament and revenue-sharing treatment respectively. Tests for equality of the median productivity for subjects opting for the fixed payment and variable payment yield continuity corrected Pearson $\chi^2(1)$ that imply p-values < 0.0001 for both productivity measures in the piece-rate treatment and the tournament treatment and p-values < 0.05 for both productivity measures in the revenue-sharing treatment.

¹⁶The Spearman rank correlations and corresponding p-values for verbal IQ test scores of German participants and Productivity Indicators 1 to 3 are respectively: -0.127 (p-value < 0.05), -0.119 (p-value < 0.05), 0.114 (p-value < 0.06). The Spearman rank correlations and corresponding p-values of math grades and Productivity Indicators 1 to 3 are respectively: -0.296 (p-value < 0.001), -0.198 (p-value < 0.002), and 0.286 (p-value < 0.001). The Spearman rank correlations and corresponding p-values of *Abitur* grades and Productivity Indicators 1 to 3 are respectively: 0.267 (p-value < 0.001), 0.136 (p-value < 0.011), and -0.224 (p-value < 0.001).

The sorting pattern is similar in all three treatments (see Panel (b) for the tournament and 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.

RISK ATTITUDES, RELATIVE SELF-ASSESSMENT, SOCIAL PREFERENCES AND GENDER

In the previous section we have shown that productivity systematically affects self-selection into different incentive schemes. In this section we study further potential drivers of self-selection: risk attitudes, relative self-assessment, social preferences and gender. Table 1 provides raw correlations of these variables. We use a multiple regression approach to determine how these factors jointly affect the self-selection decision; and to understand their relative importance. In presenting our results we mostly refer to Table 2, which reports estimates of marginal effects from Probit models in which the latent variable is the propensity to opt for the variable pay alternative. Columns (1) to (3) contain the results for the piece rate treatment, the tournament treatment and the revenue-sharing treatment respectively. Column (4) displays results for the whole sample, pooling all decisions for variable vs. fixed payments. The set of explanatory variables consists of risk attitudes, relative self-assessment, social preferences and gender, in addition to productivity. In line with the results above, Table 2 shows that productivity is significant in all specifications. The respective coefficient in Column (1), e.g., implies that solving one additional problem in the five minute working task (Productivity Indicator 3) increases the likelihood of selecting into the piece rate by 4.4 percent.

Risk attitudes are an obvious candidate to affect the sorting decision into different incentive schemes.¹⁷ Risk is always involved in variable payment schemes simply because incomes are uncertain. In contrast no risk is involved in the fixed wage payment since payments are independent of output. As a consequence the expected utility from variable pay is lower for risk averse subjects than for risk neutral or risk loving subjects. Hence we expect that subjects are less likely to select into variable pay the more risk averse they are. This effect should be particularly strong in the tournament treatment since earnings uncertainty is most pronounced in this condition for two reasons: the spread of potential earnings is higher (0 vs. 1300 points) and the contestant's ability is not known.

In Table 2 we measure risk preferences by subjects' responses to the risk question elicited in step 12.¹⁸ It turns out that risk attitudes affect the sorting decision. Risk averse workers are less likely to self-select into piece rates (Column (1)) and tournaments (Column (2)), while no such effect is observed for the revenue-sharing treatment.¹⁹ In the piece-rate treatment, a one point higher indication of willingness to take risks on the eleven-point scale makes a subject 5.3 percent more likely to opt for the piece-rate

¹⁷Principal-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).

¹⁸We prefer this risk measure over the lottery measure elicited in step 11 since several subjects did not have a unique switching point and it is not clear how these observations should be treated. Moreover, like Dohmen et al. (2009), who have shown that answers to this question reliably predicted lottery choices in a paid field experiment, we also find a strong correlation between subjects' answers to the risk question and the lottery choices in our experiment.

¹⁹An interesting finding is, however, that among those who opt for the fixed payment in the revenue-sharing treatment, the group of subjects with above median productivity is on average 1.45 points less willing to take risks than the subjects with below median productivity. This difference is statistically significant (t-test, p-value < 0.025).

contract for a given level of productivity. The respective probability is 8.7 percent in the tournament treatment indicating that quantitatively risk attitudes matter most in this treatment. Pooling all observations from all conditions (Column (4)) yields a significant coefficient of 5.4 percent.

A fundamental difference between piece rates and tournaments is that in piece-rate schemes 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 this treatment, but could affect the sorting decision in the tournament. We would therefore expect that subjects' beliefs about their relative rank should affect the sorting decision in the tournament treatment while no such effect is expected in the piece-rate treatment.

We elicited a subject's relative self-assessed rank in step 5 of the experiment. In raw correlations relative self-assessment significantly predicts sorting into the variable pay condition in all three treatments (Spearman rank correlations; p-values < 0.0001). This suggests that subjects are more likely to select into the variable pay schemes the more productive they believe they are relative to other participants. Note, however, that raw correlations might predominantly reflect productivity sorting given that self-assessed ranks and true ranks are highly correlated: The correlation between a subject's self-assessed rank and his or her true rank based on Productivity Indicator 3 is 0.695 (see Table 1). Controlling for productivity, it turns out that the impact of self-assessment on the sorting decision is significant in the tournament schemes (p-value < 0.076), but is insignificant in the piece-rate treatment (p-value < 0.835) and the revenue-sharing treatment (p-value < 0.162), as shown in Table 2. The marginal effect estimates for the tournament scheme implies that a subject with a more positive self-assessment of one rank is about 3 percent more likely to enter the tournament than a less optimistic but equally productive subject (Column (2)). In other words subjects are more likely to select into tournaments the better they think their relative performance is, for a given true rank. If a subject can be classified as overconfident if he or she overestimates his or her true rank these results also suggest that overconfident (underconfident) subjects are more (less) likely to select into tournaments.

Another potential driver of self-selection are 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 Colin Camerer, 2003?; Fehr and Gächter, 2006?; and Fehr and Klaus M. Schmidt, 2000?). The co-existence of selfish and reciprocally motivated agents changes the optimality conditions of different types of contracts. For example, Fehr, Alexander Klein and Schmidt (2007)? 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 Christian Grund and Dirk Sliwka (2005)?, Florian Englmaier and Achim Wambach (2008)?, and Kosfeld and von Siemens (2007)?. Given the relevance of social preferences for optimal contracting it is important to understand whether they also affect the sorting of agents.

Remember from Section B that all subjects participated in a trust game in both roles as first and second mover, which allows us to assess willingness to trust and reciprocal inclination for each subject. As an indicator for trust we use the amount sent to the second mover. To classify agents with respect to their reciprocal inclination, we determined for each subject the relation between transfer and back transfer running simple OLS regressions of the back transfers on received transfers, forcing the slope to go through the

origin. Results in Table 2 show that neither trust nor reciprocity significantly determine the sorting decision. This holds for all treatments with the exception of a very small effect for trust in the piece-rate treatment.

Finally we study whether gender affects the sorting decision into variable pay. In a recent paper Niederle and Vesterlund (2007)? show that women shy away from competition. This finding 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, according to our findings in Section A, translate directly into lower average wages for women than for men. Results in Table 2 suggest that gender does not significantly affect the choice between the fixed payment and the variable pay alternative in any treatment, conditional on productivity, risk attitudes, relative self-assessment and social preferences.²⁰

Before jumping to the conclusion that there are no gender differences in the sorting decision, one might raise the question whether the conditional probability is a useful measure for answering the question of whether women tend to avoid competitive environments. In fact, if there are important gender differences in attributes that we condition on and that affect the sorting decision, then gender differences in sorting choices need not be unveiled by a significant coefficient for the gender dummy, but might be reflected in significant estimates of gender specific attributes. Risk attitudes might be such an important factor through which gender differences manifest: for one thing, our findings show that risk attitudes affect the sorting decision. At the same time, accumulating evidence reveals that women tend on average to be more risk averse than men (for an overview of studies see Rachel Croson and Gneezy, 2009?). This result is confirmed by the significantly negative correlation between being female and risk attitude reported in Table 1.²¹

In order to check whether there is reason to believe that gender differences in sorting behavior is camouflaged by attributes, in which men and women differ and that affect the sorting decision, we revert to the raw data. We show in Table 4 that women seem indeed to prefer fixed wages to a higher extent than men in our experiment. The table reveals that only 46.4 percent of all women prefer the variable pay alternative compared to 69.8 percent of all men. This gender difference is very strong in the piece-rate treatment and in the tournament treatment, and somewhat smaller in the revenue-sharing treatment (albeit a 20 percentage point difference between men and women). The raw gender difference is highly statistically significant. In unreported regressions we investigated how the size and significance of the gender difference in sorting behavior change when we step-wise add controls for productivity and various preferences and attitudes. It turns out that the gender difference remains significant if we only condition on productivity. Figure 5 illustrates this result by comparing the sorting decisions of men and women 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 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

²⁰The coefficient on the gender dummy becomes statistically significant at the 10-percent level only in the tournament treatment once we control for personality differences. Here, the marginal effect implies that women are 22 percent less likely to enter a tournament.

²¹In fact, 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 4.84 for women and 5.80 for men (medians are 5 and 6, respectively). This is in line with the findings of Dohmen et al. (2009)?.

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.

In a next step, we added risk preferences. The coefficient on the gender dummy becomes insignificant and also considerably smaller in size once we control for productivity and risk preferences. Similar to Nabanita Datta Gupta, Anders Poulsen and Villeval (2005)?, but contrary to the findings by Niederle and Vesterlund (2007)?, part of the gender differences can thus be attributed to differences in productivity and risk preferences in our set-up. 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 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.

Concluding our discussion of Table 2, note that our results are robust to controlling for personality traits using a variant of the so-called Big-5 inventory.²² This is important since personality itself could be an important driver of self-selection but also since it could be correlated with our other characteristics, e.g., risk attitudes (Angela Lee Duckworth et al., 2008?).

In our analysis so far we have assumed that attitudes and preferences affect the sorting decisions of all subjects in the same way, i.e., regardless of how productive they are. An interesting possibility is that the extent to which particular personal characteristics affect the sorting decision depends on the location in the productivity distribution. In other words, personal characteristics are likely to matter differently for “marginal types”, i.e., for subjects whose selection decision is on the fence than for subjects whose decision is basically determined by their productivity (or other characteristics). To illustrate consider the role of risk attitudes in the piece-rate treatment. One would expect risk attitudes to matter most for workers who are close to the productivity threshold for a risk neutral agent: After all, very productive subjects who are far above the productivity threshold should always prefer the piece rate regardless of their risk attitudes. Likewise, very unproductive subjects should always prefer the fixed payment, again regardless of their risk attitudes. In order to test for multi-dimensional sorting among marginal and among non-marginal types we ran the same regressions as in Table 2 on particular sub-populations in Table 3. Determining marginal types is relatively straightforward in the piece-rate treatment. In Column (1) of Table 3, we estimated the model from Table 2 on the sample of 76 subjects who solved between 10 and 30 problems in step 3 of the experiment. We chose these numbers because 20 problems would be a risk-neutral subject’s productivity threshold in the piece-rate treatment if 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 if subjects expect to produce twice as much when working time is twice as long. The results reported in the first column of Table 3 confirm that estimated marginal effect of our risk measure is in fact larger in this sub-sample (7.3 percent) compared to the whole sample in Column (1) of Table 2 (5.3 percent). Thus the role of risk attitudes in the piece-rate treatment depends on the location in the productivity distribution.

Determining marginal types for the tournament and the revenue-sharing treatment based only on productivity is not feasible, since decisions are based on beliefs about other participants’ productivity and potentially other factors as well. We therefore propose a different approach to define marginal and non-marginal types based on response times.

²²See footnote 7.

The idea is that marginal types need more time to determine their decision than non-marginal types. In the piece-rate treatment, e.g., it is an easy and quick decision to select the variable pay mode if one is either very productive or very unproductive. Subjects with intermediate productivity supposedly need more time and deliberation to decide. Participants' response times are automatically recorded in each step of the experiment. For each subject we know how much time has elapsed between the presentation of the sorting decision screen and their actual decision. In Columns (2) to (7) of Table 3 we classify subjects according to median response times. A subject who needed more time to decide than the median subject is considered a marginal type; those who needed less time are classified as non-marginal types. For each treatment we ran two separate regressions one for marginal, one for non-marginal types. Columns (2) and (3) show the results for the piece-rate treatment. In support of our response time based notion of marginal types, it turns out that marginal types defined by response times show an extremely similar pattern to marginal types classified according to threshold productivity (compare Columns (1) and (2)). In particular, the coefficient for risk attitudes is quite large (marginal effect is 7.6 percent) and highly significant. This is in sharp contrast to non-marginal types, where risk attitudes play no role at all.²³ Turning to the tournament treatment we again find a differential pattern. Non-marginal types, i.e., those who decide relatively quickly base their decision on productivity and relative self-assessment. In both cases marginal effects are larger than for the whole sample (2.3 vs. 1.8 percent for productivity and 4.6 vs. 2.7 percent for relative self-assessment). The coefficient on risk attitudes is not significant. Risk attitudes matter strongly, however, for marginal types. For marginal types the marginal effect is 12.3 percent and considerably larger than for the whole sample (8.7 percent). In the revenue-sharing treatment marginal types are more likely to enter this pay scheme the more productive they are.²⁴ Taken together the results in Table 3 show that the role of personal characteristics for subjects' sorting decision depends on whether they are close to the fence, as proxied by how long they need to take a decision.

C. Effort Provision and Output Changes

Although the main focus of this paper is on sorting, it is interesting to know whether participants' performance is affected by different incentives. In this section we therefore discuss how output and effort provision vary across different incentive schemes. Intuitively, subjects in the variable pay schemes should provide at least as much effort as subjects who are paid according to a fixed-payment contract, simply because 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. Table 5 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

²³Note that response times per se are not significantly related to gender, preferences and attitudes (risk attitudes, relative self-assessment, trust, reciprocity) as linear regression analysis shows; and none of these characteristics significantly predicts the type classification (marginal vs. non-marginal) as a probit analysis reveals.

²⁴Non-marginal types are more likely to opt for revenue-sharing the more trusting they are, while non-marginal types are less likely to enter the revenue-sharing scheme the more trusting they are. One potential explanation is that more trusting marginal subjects have less optimistic beliefs about the teammate's productivity (from the 5-minute condition) rendering it more attractive to select the revenue-sharing treatment *ceteris paribus*.

(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 5.60 for those who later selected into the piece rate while it is 5.49 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 completely, 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, for example, mean effort is now 6.03 for workers receiving a piece rate, while it is only 4.26 for the fixed wage group. This difference is highly statistically significant as the p-value in Column (6) shows. Results in Table 5 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, also when controlling for individual productivity.²⁵ A comparison of efforts in Columns (2) and (5) further shows 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.²⁶ A comparison of Columns (1) and (4) reveals that self-reported effort for the 10-minute work task is higher in all treatments. This increase is highest for tournament participants and lowest for participants in the revenue-sharing scheme. Table 5 also indicates that subjects who work for variable pay tend to get more exhausted, but differences in exhaustion levels are never significant.

Our data also permit us to shed some light on the effort response to variable pay (i.e., the incentive effect) by comparing within-subject output differences between the 5-minute work task and the selected variable payment conditions for participants who sorted into the variable pay schemes. To test, for example, whether tournament incentives trigger larger changes in output than piece-rate or revenue-sharing incentives – as suggested by the self-reported measures – we compare output in the 5-minute work period with output in the first 5 minutes of the 10-minute work period for participants in these three different variable incentive schemes. We use output differences within subjects to control for ability differences and consider only output of the first 5 minutes of the 10-minute task because of potential fatigue effects. Output is statistically significantly higher in the first five minutes of the 10-minute work task than in the 5-minute work task in all treatments as Wilcoxon tests reveal (piece-rate treatment: p-values < 0.0051; tournament treatment: p-value < 0.0003; revenue-sharing treatment: p-value < 0.031). The average increase in output is highest in the tournament treatment (3.4 more answers, t-test, p-value < 0.002) followed by the piece-rate (1.8, p-value < 0.007) and revenue-sharing treatment (1.5, p-value < 0.035).

²⁵The results are available upon request.

²⁶In this respect it is interesting to compare effort differences within the sample of subjects who selected into the fixed payment. In a regression in which the dependent variable takes the value one if subjects in the fixed payment scheme provide low output (output between 0 and 10, 8 percent of subjects in the fixed payment) on the same set of control variables as in Table 2, we do not find any statistically significant effects (at the 5-percent level), with the exception of productivity (i.e., indicating that more productive workers are less likely to produce low output).

III. Discussion

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 holds for piece-rate, tournament as well as revenue-sharing schemes. Incentives also affect effort provision. In our study workers report higher effort levels in pay for performance schemes than in fixed payment schemes. Moreover, they report higher levels of stress and exhaustion. Yet, 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 piece-rate or tournaments and fixed wages. Risk attitudes do not seem to matter at all for the decision to select into team incentive schemes. Tournaments attract workers who believe their performance is high, relative to other workers. This effect plays no role in piece-rates, which makes sense because in this treatment payoffs are independent of the output of others. Variable payment schemes generally attract fewer women, an effect that is partly driven by an underlying gender difference in risk attitudes and productivity. The effect is strongest in the most competitive scheme, the tournament. Additional results show that the impact of personal characteristics on the sorting decision is heterogenous, and that social preferences seem to play only a marginal role. Sorting generally depends on a subject's location in the productivity distribution or on whether a subject's decision is on the fence.

While productivity is a strong and significant determinant of sorting into all variable pay schemes, the importance of preferences and attitudes on the sorting decision depends on the type of variable incentives. On the whole, we find less robust effects of preferences and attitudes on sorting into our revenue-sharing scheme compared to sorting into the piece-rate or tournament scheme. We interpret this finding to reflect that the decision between revenue-sharing and fixed wages is quite complex and definitely more difficult than, for example, between a fixed payment and a piece rate. Support for this interpretation comes from the fact that subjects need significantly longer to decide than in the piece rate treatment. Respective median response times for the piece-rate, tournament, and revenue-sharing treatment are 34.5 seconds, 72 seconds, and 116 seconds, respectively. In addition variance in output is significantly higher in the revenue-sharing treatment suggesting that not only the decision to enter is complex but also that conditional on having selected into the revenue-sharing treatment the incentive structure is more complex than in the other two treatments.

It is interesting to know whether the observed sorting has led to an efficient outcome. In other words one could ask whether subjects made the "right" sorting decision and how this depends on the incentive scheme. Deriving individual optimality conditions is not straightforward, however, given that we use a real effort task and therefore do not know individual effort costs, and given that preferences and attitudes are relevant for the evaluation of particular sorting outcomes. What can be done, however, is to check whether subjects behave in a revenue maximizing way. This is relatively straightforward for subjects who opt for the variable pay as their counterfactual earnings (400 points) are known. However, for subjects who opt for the fixed payment, it is difficult to assess whether their choice was revenue maximizing as the counterfactual earnings in the variable pay scheme are not observed. We therefore concentrate on analyzing whether the sorting decisions of those who opted for the variable pay are revenue maximizing.

In the piece rate treatment, 64 of the 73 individuals (87.7 percent) who opted for the piece rate earned more than 400 points, and 9 individuals earned less. For all but one of the nine individuals who earned less than 400 points in the piece rate, the sorting decision was also not optimal ex ante if these subjects expected to solve twice as many

problems in 10 minutes than they solved in 5 minutes. In the tournament treatment, we can ex post assess whether choices of tournament participants were revenue-maximizing in the following way: Given the output of a subject who chose the tournament we can check how many other subjects who have opted for the tournament had a higher or a lower output. This determines the expected probability of winning given the realized sorting pattern. Multiplying this probability with the winner prize in the tournament (1300) yields an expected payoff. If this expected payoff exceeds the fixed payment (400) the sorting decision was revenue maximizing. This holds for 42 of the 60 (i.e., 70 percent) tournament participants. In the revenue-sharing treatment, 60 of the 76 subjects (78.9 percent) who opted for the revenue-sharing contract end up with a payoff of more than 400 points. This does not necessarily imply, however, that subjects who ended up with less than 400 points did not make a sorting decision that was revenue-maximizing in expectations because their decision had to be based on expected output of all participants who opted for revenue-sharing. To address this issue, we take a subject's output in revenue-sharing teams and add it to the average output of all other subjects who worked under the revenue-sharing scheme in order to calculate expected output in revenue-sharing teams, given the observed sorting and output patterns. It turns out that this expected output is smaller than 80 (i.e., expected revenue is smaller than 400) for only 7 subjects.

Even though we have already highlighted multiple dimensions of sorting, the sorting process is in general much more complex. For example, self-selection into jobs will in general not only depend on personal characteristics but also on the nature of the job and task. In this paper we have studied one particular type of task while varying incentives; it would also be important to study sorting dependent on the nature of the task for a given incentive, for example tasks that differ with respect to self determination, intrinsic interest, how challenging it is, etc.

An interesting question is whether the sorting patterns that we observe in the lab generalize qualitatively to labor markets outside the lab. Are individual characteristics such as productivity or risk attitudes systematically correlated with modes of payment? We can investigate this issue for the German labor market by analyzing data from the German Socio-Economic Panel Study (SOEP). The SOEP is a representative panel survey of the adult population living in Germany. All members over the age of 17 of a household in the sample are asked for a wide range of personal and household information, and for their attitudes on assorted topics.²⁷ Each wave records information on the respondents current labor market status. The 2004 wave of the SOEP included an additional question on whether the performance of a respondent is regularly evaluated in a formal procedure, a requisite element of performance contingent remuneration schemes (see Grund and Sliwka, 2009?). As a proxy for productivity we use years of education, experience in full-time and part-time employment, and tenure. Respondents in the 2004 wave also answered a survey question on their willingness to take risks that subjects in the experiment answered and that has been experimentally validated in Dohmen et al. (2009) (See section B). In addition, we have survey measures on individuals' trust attitudes from the 2003 wave and measures of reciprocity from the 2005 wave.²⁸ We also control for age.

We estimate Probit models in order to assess how the probability of working under a variable payment scheme is related to productivity, risk attitudes, social preferences and gender. Table 6 reports marginal effect estimates. The dependent variable in both specifications takes the value 1 if the respondent's performance is regularly assessed in a formal evaluation procedure. This measure is a rough indicator of variable pay; and in

²⁷For more details on the SOEP, see www.diw.de/gsoep/.

²⁸For further details, see Dohmen et al. (2008)?.

a sense corresponds to our pooled regression, shown in Column (4) of Table 2 where we combine information from different variable incentive schemes.

Column (1) reveals that the probability to work under some variable pay scheme increases in productivity, measured in terms of years of schooling, experience and tenure. For example, one year of schooling increases the probability by 2.2 percentage points. Consistent with the sorting decisions in the experiment we find that more risk tolerant workers are more likely to work in jobs with performance evaluation. The effect is significant both statistically as well as economically. The marginal effect of 0.010 implies that a worker who is completely willing to take risks on the 11-point scale has an about 10 percent higher likelihood to work under variable pay than a worker who is completely unwilling to take risks, assuming that the marginal effect at the mean applies to the entire range of admissible answers. This corresponds to about 4.5 years of schooling. The estimates also suggest that individuals who are less reciprocal are more likely to work for variable pay; the impact of trust is not significant. Women are on average 6 percent less likely to work for variable pay than men, a result that is similar to our experimental results. As in the experiment, this effect is considerably higher if we do not include controls for risk attitudes and productivity. The coefficient estimates in Column (2) show that these findings are largely robust to controlling for firm size, industry sector, and broad indicators for occupational categories. The latter capture the job type (e.g., white-collar vs. blue-collar) and level, ranging from untrained worker to master craftsmen for blue-collar workers and from employee with simple duties without training/education certificate to employee with extensive managerial duties (e.g., managing director, manager, head of a large firm or concern) for white-collar workers. Not surprisingly, the coefficients on productivity measures become smaller in size since the controls for occupational categories essentially capture heterogeneity in productivity. The impact of risk attitudes and gender also remains strong.²⁹

Additional complementary evidence is provided by Holger Bonin et al. (2007)[?], who use data from the SOEP and find that individuals who are more willing to take risks are more likely to work in occupations with higher earnings variability, and Dohmen et al. (2005)[?], who observe that risk averse workers are more likely to be employed in the public sector, where – according to the findings in Column 2 of Table 6 – performance pay is less prevalent. 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. An inspection of the SOEP data shows that women in Germany are more likely to work in the public sector than men. Among full-time employees, 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.

In sum these results are quite similar to our findings from the lab. Of course one should interpret the results from the field with care: unlike with the experimental data we cannot, e.g., exclude the possibility of reversed causality. Nevertheless we think that exploiting complementarities between controlled lab and large-scale field data is useful and that finding similar sorting patterns with these complementary data sets is suggestive for the importance and systematics of multi-dimensional sorting.

Multi-dimensional sorting has several important implications. When designing 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 differ-

²⁹The results also hold when we restrict our sample to the private sector only. In fact, the impact of productivity and gender even becomes somewhat stronger.

ent career paths to get the right people on the right job. For example, there might be jobs within the firm where it is important to prevent workers from excessive risk taking (e.g., safety officers), i.e., where it is important to employ risk averse workers (high downside risk for the firm). This may render variable pay suboptimal for this job. In other jobs of the same firm risk tolerance may be advantageous, rendering variable pay optimal. As a consequence the firm offers different contracts for different jobs in order to induce workers to self-select optimally. Note that in our experiment we study an extreme case: either no variable pay (fixed wages) or only variable pay (zero fixed wage component). Contracts in firms typically combine both elements. By adjusting the relative weight of variable pay, the firm can influence the allocation of workers into different jobs.

Our results also shed light on the question why firms use different incentive schemes even when operating in similar environments. A possible explanation is that they have different requirements 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

TABLE 1—CORRELATION AMONG INDEPENDENT VARIABLES

	Productivity indicator 3	Risk attitude	Relative self-assessment	Trust (amount sent)	Reciprocity	1 if female
Productivity indicator 3	1.0000					
Risk attitude	0.0874	1.0000				
Relative self-assessment	-0.6953	-0.1046	1.0000			
Trust (amount sent)	-0.0107	0.1033	-0.0110	1.0000		
Reciprocity	-0.1412	-0.0666	0.0987	0.4691	1.0000	
1 if female	-0.2644	-0.2451	0.2160	-0.1245	0.1583	1.0000

Notes: The table shows the partial correlations between the independent variables.

TABLE 2—DETERMINANTS OF SORTING

Dependent variable	1 if piece rate (1)	1 if tournament (2)	1 if revenue sharing (3)	1 if variable pay (4)
Productivity indicator 3	0.044*** [0.009]	0.018*** [0.007]	0.016*** [0.003]	0.023*** [0.004]
Risk attitude	0.053*** [0.015]	0.087*** [0.032]	0.008 [0.013]	0.054*** [0.014]
Relative self-assessment	0.003 [0.015]	-0.027* [0.015]	-0.020 [0.014]	-0.015* [0.009]
Trust (amount sent)	0.002* [0.001]	0.002 [0.002]	-0.001 [0.002]	0.001 [0.001]
Reciprocity	0.006 [0.041]	0.012 [0.098]	0.063 [0.046]	0.012 [0.034]
1 if female	0.029 [0.121]	-0.157 [0.137]	-0.097 [0.075]	-0.068 [0.059]
Pseudo R^2	0.410	0.307	0.204	0.268
Observations	120	120	120	360

Notes: Probit estimates. Marginal effects (evaluated at the mean of independent variables) reported; * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors clustered for sessions are reported in brackets below the marginal effects estimates. The smaller the value of the self-assessment variable is, the more productive a subject thinks he is relative to others.

TABLE 3—DETERMINANTS OF SORTING AT THE MARGIN

Dependent variable	Piece rate treatment 1 if piece rate chosen			Tournament treatment 1 if tournament chosen		Revenue-sharing treatment 1 if revenue sharing chosen	
	marginal type (productivity)	marginal type (response time)	non-marginal type (response time)	marginal type (response time)	non-marginal type (response time)	marginal type (response time)	non-marginal type (response time)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Productivity indicator 3	0.046*** [0.012]	0.053*** [0.012]	0.036*** [0.012]	0.013 [0.012]	0.023* [0.012]	0.023*** [0.005]	0.001 [0.005]
Risk attitude	0.073** [0.030]	0.076*** [0.027]	-0.017 [0.044]	0.139*** [0.044]	0.034 [0.051]	0.033 [0.030]	-0.007 [0.007]
Relative self-assessment	0.008 [0.018]	0.015 [0.033]	-0.020 [0.023]	-0.011 [0.022]	-0.046** [0.021]	-0.023 [0.016]	-0.023 [0.016]
Trust (amount sent)	0.001 [0.002]	0.003 [0.003]	0.001 [0.002]	0.002 [0.003]	0.003 [0.003]	-0.004*** [0.002]	0.002** [0.001]
Reciprocity	-0.001 [0.041]	0.005 [0.068]	0.022 [0.042]	-0.014 [0.098]	0.069 [0.112]	0.077 [0.050]	0.028 [0.016]
1 if female	0.001 [0.154]	0.121 [0.160]	-0.160 [0.187]	-0.086 [0.207]	-0.259 [0.197]	-0.029 [0.084]	0.011 [0.035]
Pseudo R^2	0.215	0.348	0.529	0.283	0.457	0.220	0.401
Observations	76	60	60	62	58	61	59

Notes: Probit estimates. Marginal effects (evaluated at the mean of independent variables) reported; * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors clustered for sessions are reported in brackets below the marginal effects estimates. In column (1), only observations of participants who solved between 10 and 30 problems in the 5-minute work condition correctly are considered. In columns (2)-(7) we define marginal candidates based on the response time they take to make a sorting decision. Subjects whose response time is above the median response time of the respective treatment are classified as marginal types. Subjects whose response time is below the median are classified as non-marginal types. The respective median response times for the piece-rate, tournament, and revenue-sharing treatment are 34.5 seconds, 72 seconds, and 116 seconds. 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—PROPORTIONS OF MEN AND WOMEN SORTING INTO VARIABLE PAY SCHEMES

	Piece rate	Tournament	Revenue sharing	All variable
Women	47.5	37.3	54.0	46.4
Men	73.8	62.3	73.7	69.8

Notes: The table shows the percentages of men and women who select into the variable pay schemes.

TABLE 5—EFFORT, STRESS, AND EXHAUSTION

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

	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	5.60	5.49	0.596	6.03	4.26	< 0.001
Stress	5.48	5.62	0.502	5.71	3.51	< 0.001
Exhaustion	3.05	2.74	0.317	4.07	2.68	< 0.001
Number of observations	73	47		73	47	

(b) Effort, Stress and Exhaustion in Tournament 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	5.57	5.37	0.264	6.23	4.70	< 0.001
Stress	5.43	5.48	0.524	5.80	3.95	< 0.001
Exhaustion	2.93	2.93	0.698	3.67	3.30	< 0.268
Number of observations	60	60		60	60	

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

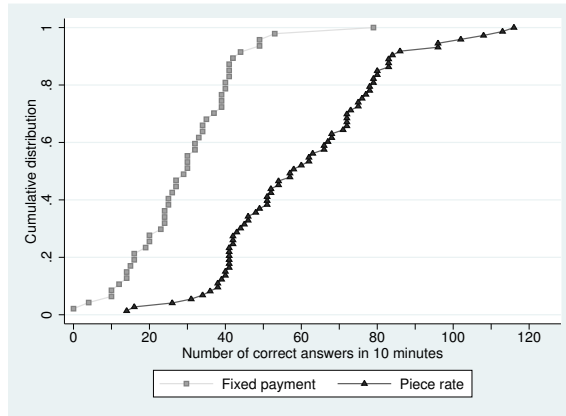
	Before sorting decision			After sorting decision		
	Revenue sharing	Fixed	M-W test	Revenue sharing	Fixed	M-W test
	(Mean)	(Mean)	(p-value)	(Mean)	(Mean)	(p-value)
	(1)	(2)	(3)	(4)	(5)	(6)
Effort	5.54	5.39	0.536	5.66	4.48	< 0.001
Stress	5.41	5.61	0.241	5.37	3.91	< 0.001
Exhaustion	2.57	2.50	0.806	3.63	2.84	< 0.020
Number of observations	76	44		76	44	

TABLE 6—SURVEY EVIDENCE ON SORTING INTO PERFORMANCE PAY

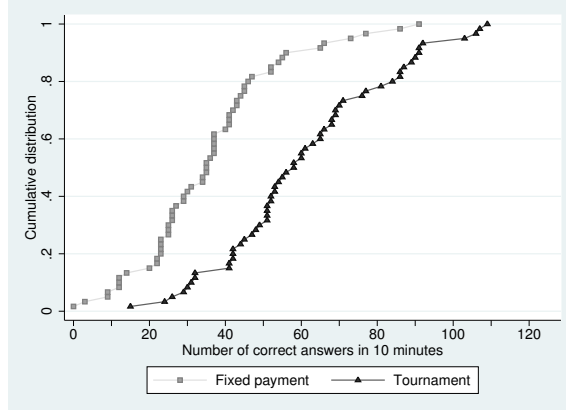
Dependent variable:	1 if performance evaluation	
	(1)	(2)
Years of schooling	0.022*** [0.002]	-0.001 [0.003]
Experience full-time (in years)	0.011*** [0.002]	0.006*** [0.002]
Experience full-time ² /100	-0.021*** [0.004]	-0.011** [0.005]
Experience part-time (in years)	-0.002 [0.003]	0.000 [0.004]
Experience part-time ² /100	0.008 [0.013]	0.003 [0.014]
Tenure (in years)	0.004*** [0.001]	-0.002** [0.001]
Age (in years)	-0.006*** [0.001]	-0.003** [0.001]
Risk attitude	0.010*** [0.002]	0.008*** [0.003]
Trust in strangers	-0.001 [0.008]	0.004 [0.008]
Reciprocity	-0.014** [0.006]	-0.008 [0.006]
1 if female	-0.057*** [0.012]	-0.041*** [0.014]
i if in public sector		-0.050*** [0.016]
1 if living in East Germany		0.020 [0.014]
Firm size dummies	no	yes
Industry dummies	no	yes
Occupation dummies	no	yes
Pseudo R^2	0.0340	0.140
Observations	8159	8110

Notes: Probit estimates. Marginal effects (evaluated at the mean of independent variables) are reported. Robust standard errors are reported in brackets; ** significant at 5%, *** significant at 1%. The data are from the 2004 wave of the German Socio-Economic Panel Study (SOEP), except for the survey measures of trust and reciprocity which are taken from the 2003 and 2005 wave respectively. The dependent variable is a binary variable that takes the value 1 if the respondent's performance is regularly evaluated by a supervisor according to a formal procedure. The sample is restricted to respondents aged 65 and younger. We exclude respondents who are either self-employed, not in regular employment, enrolled in school or in apprenticeship.

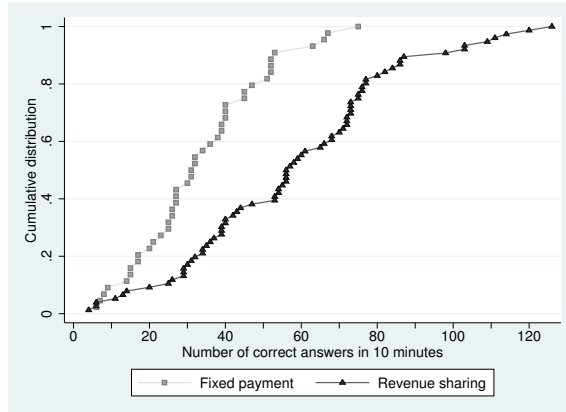
FIGURE 1. OUTPUT OF SELF-SELECTED SUBJECTS IN DIFFERENT COMPENSATION SCHEMES



(a) Piece-rate treatment



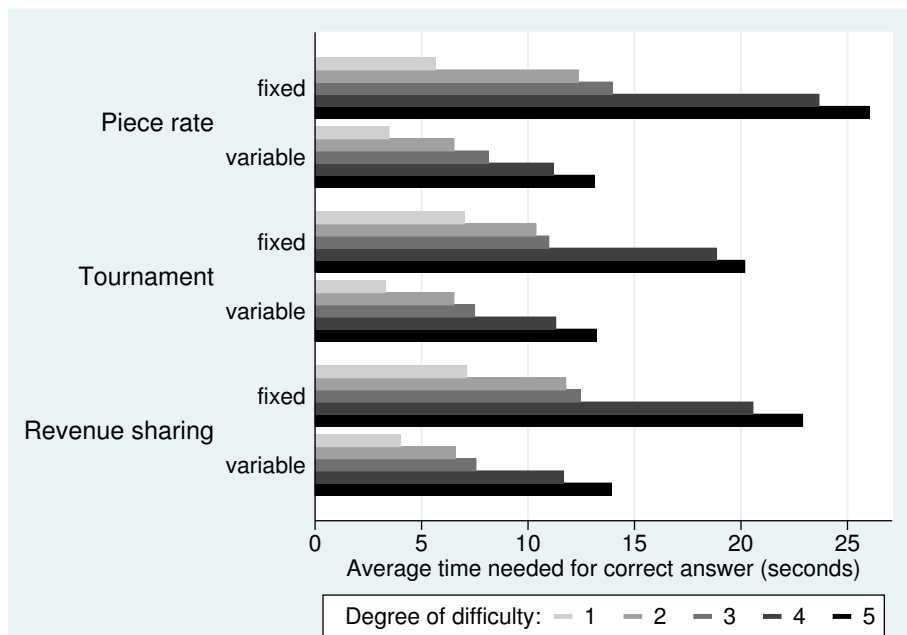
(b) Tournament treatment



(c) Revenue-sharing treatment

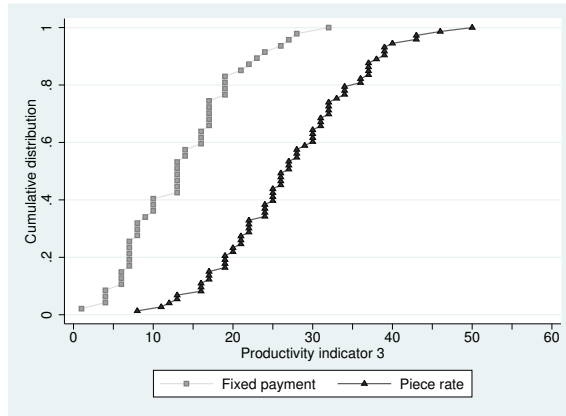
Notes: Each panel of the figure plots, for a particular treatment, the cumulative distributions of individual output (measured as the number of correct answers during the total working time of ten minutes). Panel (a) shows the cumulative output distribution in the piece-rate treatment, Panel (b) shows the cumulative output distribution in the tournament treatment, and Panel (c) shows the cumulative output distribution in the revenue-sharing treatment.

FIGURE 2. PERFORMANCE AND TASK DIFFICULTY

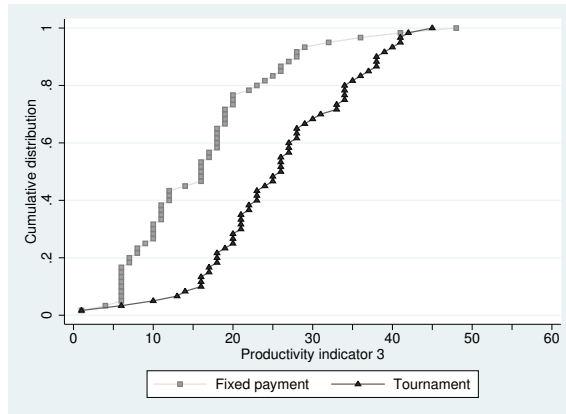


Notes: The figure shows, for each treatment, how much time (in seconds) subjects working in a particular self-selected regime need on average to solve a question of a given degree of difficulty.

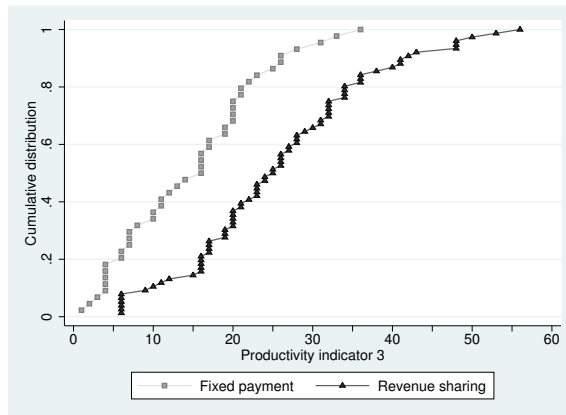
FIGURE 3. PRODUCTIVITY OF SUBJECTS BEFORE SELF-SELECTION INTO INCENTIVE CONTRACT



(a) Piece-rate treatment



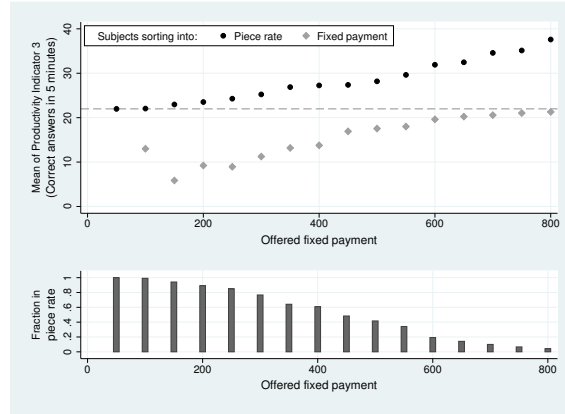
(b) Tournament treatment



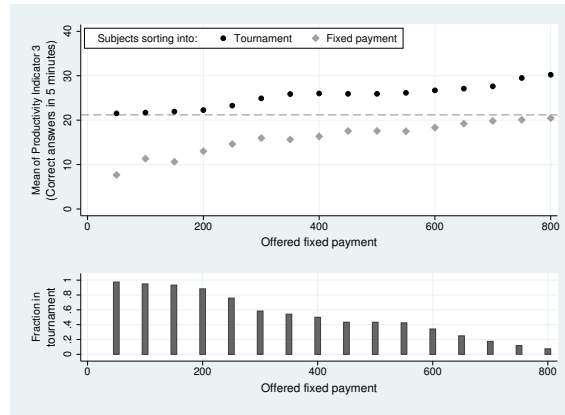
(c) Revenue-sharing treatment

Notes: Each panel of the figure plots cumulative distributions of Productivity Indicator 3, which measures the number of correct answers during a 5-minute work period, separately for subjects who subsequently chose the fixed payment alternative and those who subsequently prefer 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.

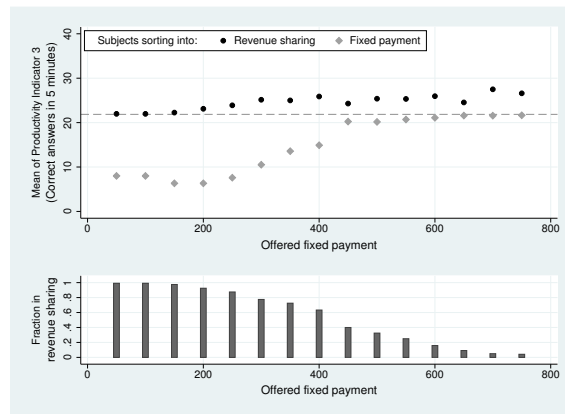
FIGURE 4. FRACTION OF SUBJECTS OPTING FOR VARIABLE PAY AND AVERAGE PRODUCTIVITY OF SORTED SUBJECTS



(a) Piece-rate treatment



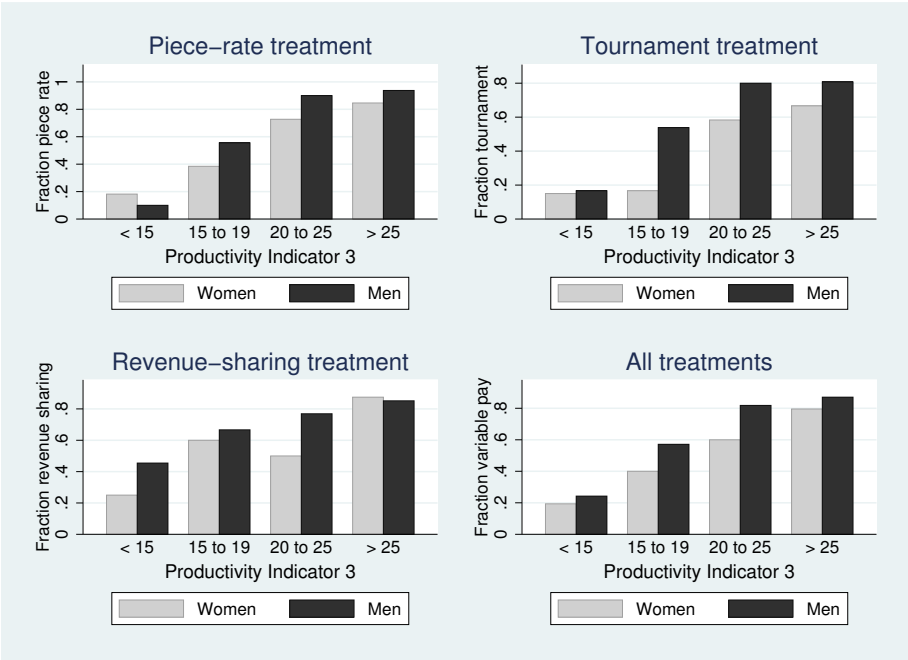
(b) Tournament treatment



(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.

Appendix

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	Questionnaires
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