Disentangling Incentive Effects from Sorting Effects: 
An Experimental Real-Effort Investigation

Dennis D. Fehrenbacher  
Institute of Business Administration, 
University of Stuttgart

Burkhard Pedell  
Institute of Business Administration, 
University of Stuttgart; 
The Wharton School, University of 
Pennsylvania

March 2012

Working Paper # 2012-08
Established in 1984, the Wharton Risk Management and Decision Processes Center develops and promotes effective corporate and public policies for low-probability events with potentially catastrophic consequences through the integration of risk assessment, and risk perception with risk management strategies. Natural disasters, technological hazards, and national and international security issues (e.g., terrorism risk insurance markets, protection of critical infrastructure, global security) are among the extreme events that are the focus of the Center’s research.

The Risk Center’s neutrality allows it to undertake large-scale projects in conjunction with other researchers and organizations in the public and private sectors. Building on the disciplines of economics, decision sciences, finance, insurance, marketing and psychology, the Center supports and undertakes field and experimental studies of risk and uncertainty to better understand how individuals and organizations make choices under conditions of risk and uncertainty. Risk Center research also investigates the effectiveness of strategies such as risk communication, information sharing, incentive systems, insurance, regulation and public-private collaborations at a national and international scale. From these findings, the Wharton Risk Center’s research team – over 50 faculty, fellows and doctoral students – is able to design new approaches to enable individuals and organizations to make better decisions regarding risk under various regulatory and market conditions.

The Center is also concerned with training leading decision makers. It actively engages multiple viewpoints, including top-level representatives from industry, government, international organizations, interest groups and academics through its research and policy publications, and through sponsored seminars, roundtables and forums.

More information is available at http://opim.wharton.upenn.edu/risk.
Disentangling Incentive Effects from Sorting Effects: An Experimental Real-Effort Investigation

Dennis D. Fehrenbacher\textsuperscript{a}, Burkhard Pedell\textsuperscript{a, b, *}

\textsuperscript{a} Institute of Business Administration, University of Stuttgart, Keplerstr. 17, 70174 Stuttgart, Germany

\textsuperscript{b} Risk Management and Decision Processes Center, The Wharton School, University of Pennsylvania, 3730 Walnut Street, Philadelphia, PA 19104, United States

* Corresponding author. tel.: +49 711 685 83170; fax: +49 711 685 83151. E-mail addresses: dennis.fehrenbacher@bwi.uni-stuttgart.de (D. D. Fehrenbacher), burkhard.pedell@bwi.uni-stuttgart.de (B. Pedell)

ABSTRACT

This study separates and compares sorting and incentive effects and analyses their contingencies in a real-effort laboratory experiment. Depending on the treatment, subjects have a choice between a piece rate, a budget-based and a fixed pay, or have no choice but are assigned to a compensation scheme. The results suggest that sorting effects have a greater impact on productivity than incentive effects. Variance in productivity within self-selected groups is smaller than within assigned groups. As to the contingencies of incentive effects, we find that anxiety and internal control conviction have a negative influence on productivity, whereas the perceived level of challenge has a positive influence. Self-selection into contracts with higher magnitude of incentives is found to be positively influenced by need for achievement and negatively influenced by internal control conviction and risk aversion. The analyses also show that skill and monetary compensation account for a large proportion of the incentive and sorting effects.

Keywords: Sorting Effect, Incentive Effect, Risk Preferences, Intrinsic Motivation, Locus of Control

JEL Classification: J33, J41, D86, J4, M52, M51
1 Introduction

Over the last two decades, a large number of firms in many countries have substantially increased their use of variable compensation schemes for work performance (cf. Institute of Management & Administration 2008). In 2010, 71 percent of global organizations and 66 percent of U.S. companies changed or planned to change their performance metrics. Additionally, 47 percent of global operations increased or planned to increase the performance thresholds of their compensation programs in 2010 (cf. Institute of Management & Administration 2010).

The figures show that organizations are frequently and increasingly using pay-for-performance programs, which tie compensation to employee productivity. These programs are primarily implemented to motivate employees to strive for superior productivity (incentive effects) and to attract adequately skilled employees (sorting effects).\(^1\) Agency theory, which is the primary economic theory used to explain when different types of monetary incentives should be used and how they should be structured, supports the argument that incentive and sorting effects are caused by monetary incentive systems (cf. Demski and Feltham 1978; Fama 1980; Jensen and Meckling 1976; Salanié 2005; Salop and Salop 1976).

However, there is very limited empirical knowledge about the interaction and relative importance of incentive and sorting effects, and about the contingencies, in particular, of sorting effects. Field studies often suffer from inability to control for potential confounding factors and inability to collect data about individual attributes. Therefore, in this paper, we employ a real-effort laboratory experiment to separate and compare incentive and sorting effects and to analyze their contingencies in a real-effort laboratory experiment. After collecting data about individual productivity, the central idea of the experiment is to divide subjects into several groups that, depending on the treatment, have a choice between a piece rate, a budget-based and a fixed pay compensation scheme, or have no choice but are assigned to one of the compensation schemes. This allows us to separate incentive and sorting effects in an integrated experiment.

---

\(^1\) Monetary incentive effects are any direct effects that monetary payments have on individual behavior. Sorting effects are the indirect effects of monetary incentives because sorting effects occur as a result of the monetary incentives (i.e., contracts) selected by employees (cf. Chiappori and Salanié 2003; Gerhart and Rynes 2003; Lazear 1986, 1998). A goal of monetary incentives that is not addressed in this study is the coordination of divisions. Groves and Loeb (1979) discuss the problems associated with coordinating and controlling divisions. Groves and Loeb (1979) also suggest a control structure that incentivizes divisional managers to transmit accurate information to the headquarters and to act in accordance with the overall company goals.
Our results show that sorting effects have a greater impact on productivity than incentive effects. Variance in productivity within self-selected groups is smaller than within assigned groups. As to the contingencies of incentive effects, we find that anxiety and internal control conviction have a negative influence on productivity, whereas the perceived level of challenge has a positive influence. Self-selection into contracts with higher magnitude of incentives is found to be positively influenced by need for achievement and negatively influenced by internal control conviction and risk aversion. The analyses also show that skill and monetary compensation account for a large proportion of the incentive and sorting effects.

Existent empirical research on incentive effects has found that monetary incentives do not always improve individual task performance. Some studies find strong positive relationships between incentives and performance (cf. Cadsby et al. 2007; van Dijk et al. 2001; Waller and Bishop 1990; Wright and Aboul-Ezz 1988), whereas others find that incentives have no effect on task performance (cf. Awashti and Pratt 1990; Pokorny 2008). Some scholars even find evidence suggesting that monetary incentives might have detrimental effects on performance (cf. Ashton 1990; Gächter et al. 2001; Gneezy and Rustichini 2000). There is widespread agreement in the literature that incentive effects depend on contingencies. Studies examining incentive issues report that environmental, individual and task contingencies interact with compensation schemes to determine employee effort and productivity levels (cf. Bonner 2008; Bonner and Sprinkle 2002; Bonner et al. 2000; Camerer and Hogarth 1999; Prendergast 1999; Sprinkle and Williamson 2007).

With regard to the sorting effects of incentive schemes, scholars generally agree that pay schemes with relatively high performance-dependent components tend to attract employees who are confident in their abilities to achieve superior results and tend to repel employees who have lower than average productivity levels. Thus, if employees are free to choose among various incentive schemes, then one can expect certain employees to be matched with certain incentive schemes. This matching process has been termed ‘sorting’ or ‘self-selection’ in the accounting and economic literature (cf. Chiappori and Salanié 2003; Demski and Feltham 1978; Lazear 1986, 1998).

However, recent studies focusing on the issue of sorting acknowledge that previous attempts to test sorting effects, despite their importance, have been quantitatively and qualitatively limited
compared with previous attempts to test incentive effects (cf. Cadsby et al. 2007; Eriksson et al. 2009; Eriksson and Villeval 2008). Most prior research on monetary incentives has addressed incentive effects (cf. Bonner et al. 2000; Gerhart and Rynes 2003; Lazear 2004). Those studies that consider sorting effects often only briefly address these effects or fail to account for incentive effects while analyzing sorting effects (cf. Banker et al. 2001; Burks et al. 2009; Dohmen and Falk 2011; Román 2009).

To the best of our knowledge, only Lazear (2000) provides figures comparing the sizes of incentive and sorting effects in a field experiment. His field study reports that incentive and sorting effects are equally responsible for a total productivity increase of 44 percent. Our study will pay particular attention to the question of whether incentive or sorting effects are the stronger determinants of productivity. Accordingly, our first research objective is to disentangle incentive effects from sorting effects in a work environment.

Additionally, we shall compare these effects based not only on their outputs (i.e., based on their influence on productivity levels) but also on their inputs. The input-based comparison of incentive and sorting effects involves the issue of contingencies. In addition to skill and monetary incentives, we shall study the influence of additional contingencies on these effects. Prior studies have extensively examined the incentive properties of monetary payments and thereby a variety of contingencies. Scholars agree that incentive effects do not have the same impacts across all types of individuals and all types of environments (cf. Bonner 2008; Bonner and Sprinkle 2002; Bonner et al. 2000; Camerer and Hogarth 1999; Prendergast 1999; Sprinkle and Williamson 2007).

However, previous scholars have conducted limited investigations of the contingencies of sorting effects. Therefore, we consider several contingencies in the same setting for incentive and sorting effects. This feature is an important advance on earlier studies, which have mostly regarded either incentive effects (cf. Paarsch and Shearer 2000; Paarsch and Shearer 2007; Shearer 2004) or sorting effects (cf. Dohmen and Falk 2011; Eriksson et al. 2009; Eriksson and Villeval 2008). Thus, our second research objective is to address the contingencies of incentive effects and sorting effects.

To study the objectives and to control potential confounding effects, data has been gathered by means of a laboratory experiment. When testing contracting issues empirically, the control of
exogenous and endogenous variables is a methodological challenge (Ackerberg and Botticini 2002). The advantage of the underlying laboratory experiment is that the environment is purposefully built as well as manipulated and exogenous variables can be controlled to a greater extent than in field environments. In Lazear’s (2000) empirical analysis, the introduction of a new performance-dependent payment scheme followed shortly after a new management took charge. This environmental fact could have confounded his results concerning the observed productivity gains attributed to the introduction of the performance-dependent payment scheme. In the underlying experiment, the environment is isolated from field influences, such as management interventions, by studying the effects in the laboratory. The underlying experiment also allows us to control for potential confounding factors resulting from endogenous variables such as skill. Skill might be a confounding factor, because other individual contingencies might be correlated with skill, and might be almost impossible to observe in a field environment. In the underlying experiment, skill is observed and controlled for.

The paper is organized as follows. In Section 2, we describe the experimental proceedings and design. In Section 3, we introduce the hypotheses. This section is followed by a brief display of the measurements of the individual variables in Section 4. Subsequently, based on the hypotheses, we present the findings in Section 5. In Section 6, we discuss possible management implications. In Section 7, we discuss limitations and ideas for future research.

2 Experimental Proceedings and Design

Because of the lack of natural data and to control for potential confounding endogenous or exogenous effects, we collected data by conducting a laboratory experiment. We conducted the underlying study in January 2010. The study consisted of two rounds that took place at an interval of two weeks. Round 1 was subdivided into four sets; Round 2 was subdivided into three sets. We chose an anagram task for our study because it meets several demands concerning the roles of complexity, interest, skill and experience. The anagram task requires one to find a meaningful word from a set of scrambled letters. Figure 1 provides an overview of the experimental proceedings.

*Insert Figure 1 about here*
In Set 1, we briefly introduced the participants to the experiment and collected data on the participants’ demographic information and individual attributes. In Set 2, participants practiced the anagram task for three minutes to minimize the possibility of a practice effect on the subsequent set. In Set 3, participants had to solve the anagrams for ten minutes. We took great care to ensure that all of the participants started and finished at the same time. We told the participants to solve as many anagrams as possible and offered enough anagrams that nobody was able to solve all of them during the allotted time. We did not reveal that we would use the number of anagrams solved to assess the individuals’ skills in the task. Set 4 asked for participants to rate the anagram task to determine their interest and experiences with the task.

In the two-week interval between Round 1 and Round 2, we used the skill measure in a stratified random assignment procedure to assign the participants to their respective treatment groups. The randomization procedure served as an indirect experimental control. We directly controlled the experiment by manipulating the conditions of the participants with respect to two treatment variables: magnitude of incentive and freedom of choice. The magnitude of incentive was either fixed pay (Fix), piece rate pay (Pie) or budget-based pay (Bud); freedom of choice involved either the assignment (A) or the self-selection (S) condition. Consequently, the individuals had six different conditions in Round 2: AFix, APie, ABud, SFix, SPie or SBud. These conditions constituted a 2 x 3 experimental design.

Compensation schemes are required to be different in magnitude of incentive to observe expected magnitude of incentive effects. We use not only a piece rate scheme as a type of variable contract as in e.g., Lazear (2000) or Cadsby et al. (2007) but also a budget-based scheme, as budget-based contracts may induce more effort than piece rate contracts (cf. Demski and Feltham 1978). Concerns that more than three levels in the magnitude of incentive variable could allow for an even finer discrimination in effects than just three levels can be countered by alluding to empirical evidence. In a laboratory study, Mauldin (2003) initially designed seven contracts but subsequently subsumed them under three categories for structuring her data analysis, because the results of a detailed analysis were statistically and qualitatively similar. This indicates that contracts that are too similar to each other will not track different behavior. Concerns that more than two compensation schemes might complicate data interpretation were met by refining data
analysis methods from binary to multinomial analyses and by using dummy techniques. Thus, the following three different contracts were designed.

The fixed payment contract $W_F$ is defined by:

\[(1) \quad W_F = \alpha.\]

This contract remunerates $\alpha$ independent of $x_i$, which is the number of correctly solved anagrams in Set 5 (i.e., individual productivity). The piece rate contract $W_P$ pays $\beta$ per correctly solved anagram $x_i$ as follows:

\[(2) \quad W_P = \beta \cdot x_i.\]

The dichotomous budget-based contract $W_B$ pays $\gamma$ for meeting or exceeding the set budget $B$, which is measured by the number of correctly solved anagrams, and $\delta$ for not meeting $B$, as shown by the following:

\[(3) \quad W_B = \begin{cases} 
\gamma, & \text{if } x_i \geq B \\
\delta, & \text{otherwise}
\end{cases} \]

We set the compensation scheme parameters $\alpha, \beta, \gamma$ and $\delta$ such that they produced equal expected mean payments for all three contracts over all of the participants based on their skill levels in Set 3. Because we aim to investigate the effects of skill on contract selection, this condition is necessary. The dominance of expected pay under one scheme compared with the others regardless of a participant’s skill level would hamper our investigation and bias the results. To balance the compensation schemes, we determined the parameters in the two-week interval between the two experimental rounds while assuming that a normal distribution of skill existed.

Because we promised the participants that they would earn 10 Euros on average, we set $\alpha$ to 10 Euros. We set $\beta$ to 0.23 Euros because, given the skill distribution based on Set 3 in Round 1, the participants would have earned, on average, approximately 10 Euros under the piece rate scheme if each participant had been paid approximately 0.23 Euros per correctly solved anagram. We set $\gamma$ to 28 Euros, $\delta$ to 4 Euros and $B$ to 56 anagrams. We applied the following rationale: 25 percent of the participants should be able to meet the budget to set a challenging standard for the average participant and create a high variation in pay under this compensation scheme. Thus, the budget
should be 0.67 standard deviations above the mean of the skill distribution ($55.61 = 0.67 \cdot 16.89 + 44.29$).\(^2\) Figure 2 depicts the pay-productivity functions of the three contracts.

*Insert Figure 2 about here*

Comparing the two freedom-of-choice conditions allows us to disentangle the incentive and sorting effects (Objective 1) because in the assignment condition, only the incentive effects, given the differences in magnitudes of incentive, are expected to have an influence, whereas in the self-selection condition, both the incentive and sorting effects are expected to have an impact. In both conditions, we can observe the same outcome measure (i.e., the productivity measure). Thus, analyzing the differences in the productivity measure between these conditions can provide insight into the relationships and interdependencies of the two effects under study. We consider this feature to be one major advantage of the underlying setup, which aims to produce unbiased estimates of the effects.

In Set 5, the participants received feedback about the number of anagrams that they had solved in Set 3 and were either informed of the compensation scheme under which they had to work (i.e., assignment condition) or were allowed to choose one of the three compensation schemes on their own (self-selection condition). Furthermore, we asked the participants to answer several questions concerning their individual attributes. As in Set 1, we measure the individual attributes to investigate Objective 2.

Set 6 provided the participants with 10 minutes to work on additional anagram tasks. The number of anagrams solved in this set served as a measure of productivity. Based on this productivity

\[^2\] We assume a normal distribution. Chow (1983) notes that setting a budget without considering the learning effects in an experiment will allow the participants to achieve the budget more easily than intended. While presenting the same decoding task to their participants two times, Waller and Chow (1985) acknowledge that learning could have confounded their results. To decrease the probability of confounding learning effects, Shields et al. (1989) suggest using different tasks in the second round. Presenting their participants with different tasks of the same type for each round, Shields et al. (1989) report an increase in performance of 13 percent. Applying the same strategy, Hyatt and Taylor (2008) report an increase of 17.82 percent between a non-paid and an incentive session. Both Shields et al. (1989) and Hyatt and Taylor (2008) attribute the performance increases to incentives instead of learning effects and do not think that learning played a significant role in their experiments. Examining neuroscience evidence, Dohmen and Falk (2011) also only expect small learning effects in their experiment. In light of these reports, we acknowledge that learning effects impact performance levels, but we expect these effects to be small during the experiment, particularly because we used different anagram tasks in the second round. Thus, we do not adjust the set budget $B$ to account for possible learning effects.
measure, we determined the participants’ remuneration. In Set 7, the participants completed additional questionnaires concerning their individual preferences and were asked to complete a set of validity check questions. At the conclusion of the study, we asked the participants to give their final comments.

Initially, 185 participants took part in Round 1. Due to no-shows in Round 2 and some participants failing to pass validity checks, in total, the study generated 165 valid data sets. In each round, the academic sponsor of the study was present to ensure that the conditions and instructions were consistent. To minimize the waiting time and ensure smooth operations, as many as two assistants helped distribute the paper-based sets. The participants were seated such that they were not influenced or disturbed by their neighbors. They had only one set to work on at all times. We considered this condition necessary to ensure that the participants worked for only 10 minutes on the sets that involved the work tasks (Sets 3 and 6). Assuming that all of the participants worked and answered the sets under the same conditions is crucial for the data analysis and interpretation (cf. Pokorny 2008).

3 Hypotheses

We formulate the hypotheses in accordance with the two main objectives. There are two main outcome variables: productivity levels and contracts selected (i.e., magnitude of incentive in the contract selected).

3.1 Disentangling Incentive and Sorting Effects

The information structure and characteristics of the production process affect the optimal incentive contracts. Agency theory generally expects incentives to have a positive effect on performance (cf. Baiman 1990; Fama 1980; Jensen and Meckling 1976). However, scholars have found that small additional monetary incentives may lead to lower productivity levels (cf. Bonner 2008; Bonner et al. 2000; Gneezy and Rustichini 2000; Prendergast 1999; Sprinkle and Williamson 2007). Regardless, most studies and reviews agree that a substantial increase in monetary incentives is more likely to result in higher than lower productivity levels (Camerer and Hogarth 1999; Bonner and Sprinkle 2002).
Because skill is a highly important determinant of individual task performance, one must observe skill levels to obtain a realistic picture of incentives. Thus, we expect the following relationship:

**Hypothesis 1.A:** If we control for skill, then the higher the magnitude of incentive in the contract, the higher the productivity level is.

Several studies in both the lab (cf. Chow 1983; Dohmen and Falk 2011; Farh et al. 1991; Hyatt and Taylor 2008; Mauldin 2003; Shields and Waller 1988; Shields et al. 1989; Waller and Chow 1985) and the field (cf. Banker et al. 2001; Lazear 2000; Román 2009) have indicated that a relationship exists between skill and the sorting effect. Consequently, we can expect that in performance-based pay schemes, the individuals who can select their compensation schemes will outperform those individuals who are assigned to compensation schemes. We can make this claim because we expect the individuals who self-select into performance-based pay schemes to be more productive. Thus, we formulate the following hypothesis:

**Hypothesis 1.B:** Freedom of choice moderates the relationship between magnitude of incentive and productivity. Specifically, the relationship between magnitude of incentive and productivity is stronger for individuals in the self-selection condition.

Employees’ choices regarding their compensation schemes might also have an effect on productivity variance. In his field study, Lazear (2000) discovers that the variance of workers’ productivity increases when a piece rate pay scheme is introduced. He explains the phenomenon by the difference in the workers’ (mental) choices in accepting or rejecting the piece rate scheme. Whereas Lazear (2000) cannot explicitly track this choice, our setting allows us to track it in the self-selection condition. Our experimental setting does not exactly resemble Lazear’s (2000) situation because the workers observed by Lazear (2000) had been remunerated on an hourly basis before the piece rate scheme was introduced, whereas the participants in our experiment were not remunerated prior to the introduction of the incentive schemes. However, in both cases, new incentives are introduced. Thus, in the experiment, we expect that the participants’ real and mental choices lead to a higher variance in performance under the influence of incentives than in the absence of incentives:

---

3 We use the terms ‘ability’ and ‘skill’ interchangeably.
4 The evidence related to this hypothesis might show that a sorting effect exists. However, the evidence does not link individuals’ decisions with endogenous variables, such as skill. This link is established within the framework of Objective 2.
Hypothesis 1.C: The variance of performance is higher in the presence of monetary incentives than in the absence of monetary incentives.

3.2 Contingencies of Incentive Effects

In addition to magnitude of incentive and skill, we expect additional contingencies to directly influence employee productivity levels. Many scholars account for the work aversion assumption in agency theory by utilizing the concept of negative utility (disutility) of work effort. Consequently, work effort induces costs for the individual. However, one cannot make generalizations regarding work aversion across all workers and situations (cf. Herzberg et al. 1959). A number of studies show that relationships exist between the different levels of intrinsic motivation and behavioral patterns. Lowell (1952) finds for an anagram task that individuals with high achievement motives increase their output from period to period quicker than participants with low achievement motives. Priester and Petty (1995) find that people with higher needs for cognition are more likely to extensively process the information presented to them. The most recent neuroscience evidence shows that individuals’ estimates of their task performances affect their levels of neural activity and effort. Iyer et al. (2010) perform a brain-imaging study and find that participants who think they are good at the given task (i.e., optimists) show their highest levels of brain activity when they expect large gains, whereas participants who expect to perform poorly (i.e., pessimists) show highest levels of brain activity when they try to avoid losses. Thus, the situation can influence brain activity and probably influence actual behavior and productivity levels as well. As a result, we use the concept of current motivation, which involves the situation at hand, to measure intrinsic motivation as well (cf. Section 4). Rheinberg et al. (2001) show that the factors of current motivation, such as anxiety, interest and challenge, can predict performance. We formulate a non-directional expectation:

---

5 We note that some principal agency models account for the pleasure that workers may experience through work. For instance, Holmstrom and Milgrom (1991) assume that workers may find pleasure in their work up to a certain limit. Thus, incentives are required only to spur employees to exert effort beyond that limit. Workers also exert effort in fixed pay contracts.

6 Because the aim is to compare the variables that influence incentive and sorting effects, we disregard possible interactive relationships while formulating the hypotheses for Objective 2. For instance, Vecchio’s (1982) data indicate that an interactive relationship exists such that incentives result in increased performance (on a quantitative basis) for individuals with a low need for achievement, whereas he does not find evidence for this relationship among individuals with a high need for achievement.
Hypothesis 2.A: If we control for skill and magnitude of incentive, then intrinsic motivation influences productivity.

A large number of studies have suggested that people generally perceive environmental outcomes as being under their personal control or beyond their personal control and that individual differences exist regarding the perceived degree of control (cf. Dollinger and Taub 1977; Krampen 1979, 1982, 1989; Levenson 1974; Rotter 1966; Rotter et al. 1972). We call an individual who tends to believe that outcomes are the consequences of one’s own actions an ‘internally controlled’ person (‘internal’). We call an individual who believes that outcomes are unrelated to one’s own behavior an ‘externally controlled’ person (‘external’).

The findings of previous research suggest that externally controlled employees are generally less satisfied with their jobs and are less likely to be in managerial positions than internally controlled individuals (cf. Mitchell et al. 1975). The literature also suggests that participative budgeting has a positive effect on the performances of internals, whereas participative budgeting has a negative effect on those of externals (cf. Brownell 1981). Finally, studies suggest that internals tend to perceive competitive situations to be under one’s own control, irrespective of the situation’s characteristics, and attribute their performance levels to their own activities (cf. Krampen 1982).

In terms of task performance, Dollinger and Taub (1977) find that internals exert greater effort in a task than externals when the purpose for performing the task is not provided. The researchers conclude that internals are more likely to perform the task simply for the sake of completing the task. McGee and McGee (2011) suggest that internals exert more search effort than externals when the outcomes are uncertain. Ammon (2006) and Spector (1982) discuss work-related relationships and point out that internals may show higher performances in tasks that require independence and actions based on their own initiatives, whereas externals may show higher performances in tasks that are routine and that demand strict adherence to the rules and procedures of superiors.

The question at hand is whether the experimental task can be seen as diverse (i.e., a task that demands action based on the participant’s own initiative) or routine. The anagram task in our experiment requires each participant to find a meaningful word from a set of scrambled letters. Of course, solving anagrams for hours every day might seem repetitive and routine. However, the task needs to be evaluated in the given experimental context where participants did not work for
more than 10 minutes on the task at once. The task involves uncertainty in that some anagrams might be easier to solve than others. Additionally, the task is not entirely rote because finding the proper words requires not only pure processing but also memory retrieval skills. Bonner et al. (2000) classify a variety of laboratory tasks according to their level of information-processing complexity by applying a scale ranging from one to five that increases with the level of complexity. The researchers classify the anagram-unscrambling task as a problem-solving task (level five) (i.e., as one of the most complex tasks).\(^7\) Thus, we expect the following:

**Hypothesis 2.B:** If we control for skill and magnitude of incentive, then the higher the internal control preference, the higher the productivity level is.

### 3.3 Contingencies of Sorting Effects

We indirectly addressed the possibility of a sorting effect by adopting skill as a covariate. In this section, we directly address the choices regarding the incentives in the contract. We consider not only skill but also other contingencies to explain the participants’ self-selection decisions.

Demski and Feltham (1978) indicate in their agency analysis that budget-based compensation may not only motivate employees to exert more effort but also help to prevent lower-skilled workers from entering a firm because such individuals are less likely to acquire the benefits from meeting a budget. Several empirical studies support a positive relationship between skill and the selection of contracts with higher performance-dependent incentives (cf. Barro and Beaulieu 2003; Bellemare and Shearer 2006; Chow 1983; Dohmen and Falk 2011; Eriksson et al. 2009; Lazear 2000; Mauldin 2003; Shields and Waller 1988; Shields et al. 1989; Waller and Chow 1985). Thus, we formulate the following hypothesis:

**Hypothesis 3.A:** The higher the skill level is, the higher the magnitude of incentive in the selected contract.

By investigating intrinsic motivation and the sorting effect, we focus on explaining how intrinsic motivation influences sorting (Delfgaauw and Dur 2007; Georgellis et al. 2011) rather than on how extrinsic rewards affect pro-social behavior (Ariely et al. 2009; Bénabou and Tirole 2006; Holmås et al. 2010). The evidence regarding the relationship between intrinsic motivation and sorting is inconclusive. Delfgaauw and Dur (2007) suggest that higher monetary incentives reduce the quality of workers because less motivated workers apply to work at the company.

---

\(^7\) According to Bonner et al. (2000) and Hyatt and Taylor (2008), most tasks in the underlying research area consist of decoding letters, which can be classified as production and clerical tasks (i.e., level three complexity).
Studying the British labor market, Georgellis et al. (2011) find that intrinsic rather than extrinsic rewards attract individuals to the public sector and that higher extrinsic rewards deter intrinsically motivated individuals from working in the public sector. However, their study is highly context-specific, and particular findings only hold for the education or health sectors.

Another line of thought analyzes targets that are influenced by the level of need for achievement (cf. Atkinson and Litwin 1960). Need for achievement is a construct of intrinsic motivation and influences targets in terms of the task difficulty that individuals set for themselves. It can influence the magnitude of incentive in the selected contract because the contracts implicitly involve different targets. Considering that skill is expected to be the most important determinant of selection behavior, and that both the need for achievement and the need for cognition are used to operationalize intrinsic motivation, we formulate the following hypothesis:

**Hypothesis 3.B:** If we control for skill, then the higher the intrinsic motivation is, the higher the magnitude of incentive in the selected contract.

We expect the locus of control to not only influence productivity but also decision behavior. The results of prior research suggest that the institutional setting is linked to control preferences and behavior (Hyatt and Prawitt 2001; Jurkun 1978; Leblanc and Tolor 1972).

In the present experiment, the institutional setting is strongly affected by the compensation scheme. This scheme is also one element of the institutional environment in Jurkun’s (1978) study and is particularly different for the low qualified, blue-collar workers (piece rate) in comparison with the other groups under his observation. He finds that low qualified, blue-collar workers are more externally controlled than all of the other worker groups in his study. Therefore, a relationship may exist between the preferences regarding locus of control and the preferences regarding the variability of compensation such that externally controlled individuals will prefer a variable pay plan. There is also evidence that internals try to maintain their organizational independence and are less likely to subordinate themselves to authorities than externals (cf. Ammon 2006; Spector 1982). As discussed in the agency literature, incentives may align the goals of a principal and an agent (cf. Fama 1980; Jensen and Meckling 1976).

---

8 We note that Atkinson and Litwin (1960) also find that the relationship between the need for achievement and task difficulty is not continuous. They observe that more low-need-for-achievement than high-need-for-achievement individuals choose tasks which are very difficult to achieve.

9 Jurkun (1978) describes several other factors related to the differences in locus of control among specific worker groups. In our artificial experimental environment, we do not consider these factors.
Incentives can direct an agent’s behavior in a specific manner. Furthermore, in the case of an employer–employee relationship, incentives can be used to tie an agent’s actions to a superior’s goal. Thus, one can view an agent’s freedom and independence of actions as being bound by incentives and an agent might demand what we call a bondage premium, (i.e., a premium for giving up freedom because one’s actions are bound by incentives that are tied to certain actions). Because internals prefer independence and are effectively unbound in a fixed pay contract, but bound by the budget in a budget-based contract or by piece rates, we can expect an individual’s preferences regarding the locus of control to be a decisive element in the individual’s choice of incentives. Thus, we formulate the following:

**Hypothesis 3.C:** If we control for skill, then the higher the internal control preference is, the lower the magnitude of incentive in the selected contract.

Depending on individuals’ level of risk aversion, they may prefer different types of contracts. Highly risk-averse individuals might prefer fixed pay contracts, whereas risk-seeking individuals might be better served by variable contracts (cf. Baiman 1990; Demski and Feltham 1978). Previous empirical studies on contracting issues have already considered risk. However, these investigations of risk have yielded mixed findings (cf. Chow 1983; Hyatt and Taylor 2008; Waller and Chow 1985). Chow (1983) finds no evidence for the influence of risk preferences on compensation choice. Waller and Chow (1985) find that the correlation between skill and the selected performance incentives is higher when a controllability filter is present (i.e., when the participants have a higher amount of control over the environment). The researchers presume that this effect is due to the role of risk aversion. Ackerberg and Botticini (2002) show that riskier contracts are associated with less risk-averse agents. Furthermore, Dohmen and Falk’s (2011) results support the argument that risk aversion plays a role in individuals’ decisions between fixed and variable pay contracts. Giving particular weight to analytical considerations, we expect the following:10

---

10 In this relationship between risk and contract preference, one must consider the influence of skill, as higher-skilled individuals might have different preferences than lower-skilled participants even though the two groups might possess the same risk preferences because skill might influence the expectancy values in variable pay schemes and risk taking. By accounting for this interactive relationship, the hypothesis controls for the influence of skill on contract selection.
Hypothesis 3.D: If we control for skill, then the higher the risk aversion is, the lower the magnitude of incentive in the selected contract.

3.4 Modeling Thresholds

We shall illustrate the self-selection process by alluding to a threshold model. The model extends Dohmen and Falk’s (2011) model by adding the budget-based contract and by introducing additional individual preferences. The model explains the choices by utilizing three productivity thresholds: $\pi^1$, $\pi^2$ and $\pi^3$. The first, $\pi^1$ refers to the choice between the fixed pay contract $W_F$ and the piece rate pay contract $W_P$. The second, $\pi^2$ refers to the choice between the piece rate pay contract $W_P$ and the budget-based pay contract $W_B$. The third, $\pi^3$ refers to the choice between the fixed pay contract $W_F$ and the budget-based pay contract $W_B$. The purpose is to show how intrinsic motivation, locus of control and risk preference influence the productivity thresholds at which a participant is indecisive between the respective contracts.\textsuperscript{11} If an individual’s expected productivity level is above a threshold, then he/she decides to select the contract with higher monetary incentives. We will empirically test the model in the findings section.

According to agency theory, individuals are utility-maximizing and work-averse. In the model, we endow the individuals with a utility function $U(w, e, l, a) = U(w - C(e, l, a))$, where wage $w$ positively affects utility. Furthermore, we assume that individual productivity $x_i$ is dependent on $i$’s skill $\theta_i$ and that effort $e_i$ is given by the production function $x_i = X(\theta_i, e_i)$. Thus, skill influences the level of effort necessary to achieve a certain productivity level. $l_i$ denotes an individual’s internal control preference, and $a_i$ denotes an individual’s level of risk aversion. The three thresholds are modeled by the following:

\begin{align*}
(4) \quad \pi^1 (\theta, e, l, a) &= \frac{\alpha + c_P(e_{l}, l, a) - c_P(e^*_l, l, a)}{\beta}, \\
(5) \quad \pi^2 (\theta, e, l, a) &= \frac{\beta + c_B(e_{l}, l, a) - c_P(e^*_l, l, a) + B}{\gamma}, \\
(6) \quad \pi^3 (\theta, e, l, a) &= \frac{\alpha + c_B(e_{l}, l, a) - c_P(e^*_l, l, a) + B}{\gamma + \delta}.
\end{align*}

\textsuperscript{11} When the participants face the decision, they base their decisions on their expected productivity levels.
The level of effort $e_l$ negatively affects utility because effort costs $C(e)$ are positive at all levels of $e$ ($C > 0 \forall e$) and have positive and increasing returns to scale ($\frac{\partial C}{\partial e} > 0$, $\frac{\partial^2 C}{\partial e^2} > 0$). Additionally, $e \geq 0$, and at a certain level of effort $\bar{e}$, the three different contracts employed in the experiment induce the same disutility: $C_F(\bar{e}) = C_P(\bar{e}) = C_B(\bar{e})$. $e^*_P, e^*_B$ and $e^*_F$ denote the effort choices, which maximize the respondent’s utility in the fixed, piece rate and budget-based pay schemes, respectively. For instance, the term $C_P(e^*_P) - C_F(e^*_P)$ indicates the additional disutility that is produced by an increase in effort from $e^*_P$ to $e^*_F$. If the cost of the level of effort necessary to maximize the participant’s utility in the piece rate scheme increases because of, for example, a lack of ability, then the term and, hence, the threshold $\pi^1$ increases. Increasing the fixed pay $\alpha$ also increases the threshold at which a participant is indifferent between the fixed and piece rate pay plans. Increasing the piece rate pay $\beta$ decreases the threshold and thereby increases the attractiveness of the piece rate contract. Hence, if we temporarily disregard $l$ and $\alpha$, then the productivity threshold $\pi^1$ positively depends on $\alpha$ and $C_P(e^*_P)$ but negatively depends on $\beta$ and $C_F(e^*_F)$. Additionally, the thresholds $\pi^2$ and $\pi^3$ are dependent on $\gamma$ and $\delta$, respectively. Moreover, the higher the budget $B$, the higher the thresholds are.

**Intrinsic motivation:** We assume that individuals with high levels of intrinsic motivation ($IM_H$) have cost functions with lower returns to scale than individuals with low levels of intrinsic motivation ($IM_L$) such that $\frac{\partial C_{IM_H}}{\partial e} < \frac{\partial C_{IM_L}}{\partial e}$. Thus, an increase in effort generally leads to lower disutility for $IM_H$ individuals than for $IM_L$ individuals. Additionally, the disutility that is produced by an increase in effort from one contract to the other decreases such that $\pi^1$, $\pi^2$ and $\pi^3$ all decrease. Thus, lower-skilled individuals might exceed the thresholds and might opt for a contract with a higher magnitude of incentive. Of course, the effect depends on the specific intrinsic motivation construct.

**Locus of control:** We assume that $C > 0 \forall l$ and that $\frac{\partial C}{\partial l} > 0$. At any level of $l$ and for certain levels of $\bar{e}$ and $\bar{a}$, $C_B(\bar{e}, l, \bar{a}) > C_P(\bar{e}, l, \bar{a}) > C_F(\bar{e}, l, \bar{a})$ and $\frac{\partial C_B}{\partial l} > \frac{\partial C_P}{\partial l} > \frac{\partial C_F}{\partial l}$ because given the binding nature of incentives, the higher the magnitude of incentive in a contract, the higher the

---

12 In $C_F(e^*, \bar{e}, \bar{a})$ of the underlying experiment, $e^*$ theoretically captures the effort of staying in the classroom and remaining inactive during the 10-minute work period.
disutility caused by internal control preferences. For instance, if we compare piece rate pay and fixed pay plans, then we find that the higher an individual’s internal control preference is, the higher the spread \( C_p(\bar{e}, l, \bar{a}) - C_f(\bar{e}, l, \bar{a}) \). In other words, the disutility resulting from a particular internal control preference is higher for the piece rate contract than for the fixed pay contract. Additionally, the higher an individual’s internal control preference, the higher the spread is. As discussed previously, the higher the spread, the higher the threshold \( \pi^1 \) at which an individual is indifferent between the piece rate and fixed pay plans. This finding also holds for the budget-based vs. piece rate pay threshold \( \pi^2 \) and the budget-based vs. fixed pay threshold \( \pi^3 \). Consequently, \( \pi^1 \), \( \pi^2 \) and \( \pi^3 \) increase with higher levels of internal control preference. Thus, because fewer individuals exceed the thresholds, more individuals select the contracts with lower incentives.

**Risk preference:** We suppose that \( C > 0 \ \forall \ a \), that \( \frac{\partial C}{\partial a} > 0 \) and that at any level of \( a \) and for certain levels \( \bar{e} \) and \( \bar{l} \), \( C_B(\bar{e}, \bar{l}, a) > C_P(\bar{e}, \bar{l}, a) > C_F(\bar{e}, \bar{l}, a) \) and \( \frac{\partial C_B}{\partial a} > \frac{\partial C_P}{\partial a} > \frac{\partial C_F}{\partial a} \). For instance, the disutility resulting from a particular risk aversion level is higher for the piece rate contract than for the fixed pay contract. The more risk-averse an individual is, the higher the spread. Again, the higher the spread, the higher the productivity threshold at which an individual changes his/her contract selection. Hence, with higher risk aversion, \( \pi^1 \) increases, and individuals need to possess higher levels of expected productivity to select the contract with higher incentives. The same rationale applies to \( \pi^2 \) and \( \pi^3 \).

### 4 Measurement of Individual Attributes and Data Analysis

We primarily measure the individual attributes of Sets 1, 5 and 7 (cf. Figure 1) based on inventories drawn from the literature and test these attributes by conducting reliability analysis (cf. Table 5 in the Appendix). To measure intrinsic motivation, we use the concepts of need for achievement (cf. Mehrabian 1969), need for cognition (i.e., the tendency to pursue and enjoy effortful thinking, cf. Cacioppo and Petty 1982) and current motivation (cf. Rheinberg et al. 2001). We measure the need for achievement by utilizing Mikula et al.’s (1976) scale (denoted: NFA), which is based on Mehrabian’s (1969) English measure of the tendency to achieve. We measure the need for cognition using Bless et al.’s (1994) scale (denoted: NFC), which constitutes a translation of Cacioppo and Petty’s (1982) need for cognition scale. We use...
Rheinberg et al.’s (2001) measure to measure three current motivation factors (denoted: INTEREST, CHALLENGE and ANXIETY). To measure locus of control, we use Krampen’s (1979) locus of control inventory (denoted ICP). The ICP measures the internal control preferences (i.e., the tendency to believe that outcomes are the consequences of one’s own actions). The inventory is a translation of Levenson’s (1974) English inventory. Furthermore, we adopt Holt and Laury’s (2002) measure (denoted: RISKAV) to measure risk preference based on lottery choice.

When modeling the productivity behaviors, we use the number of anagrams solved correctly within ten minutes in Round 2 as an indicator of productivity and as a dependent variable. Thus, we can categorize the productivity variable as a quantitative variable, and we use (multiple) linear ordinary least squares regression analysis.

In modeling decision behavior, we always use the magnitude of incentive variable, which possesses three categories: fixed scheme, piece rate scheme and budget-based scheme, as the outcome predicted. We use the multinomial logistic regression paradigm to predict decision behavior because it best fits the threshold model.

5 Findings

We present the findings in accordance with the two objectives formulated in the introduction. First, we separate and compare the incentive and sorting effects. Then, we isolate the incentive effects from the sorting effects to specifically observe the influences of contingency factors on productivity levels and analyze the individual contingency factors of the decisions regarding the available contracts that lead to sorting effects.

5.1 Disentangling Incentive and Sorting Effects

H1.A: The descriptive data on the participants’ productivity levels support this hypothesis. Table 1 shows the mean productivity values and standard deviations (in parentheses) for each experimental treatment group (AFix, APie, ABud, SFix, SPie and SBud) and for the experimental dimensions (magnitude of incentive and freedom of choice). Examining the magnitude of incentive dimension alone, we find that productivity increases from 42.65 in the fixed pay condition to 52.42 in the piece rate pay condition until it reaches 64.19 in the budget-based condition. The skill levels of AFix, APie and ABud are distributed relatively equally with respect
to the mean and standard deviation.\textsuperscript{13} The skill levels in the self-selection condition continuously increase from SFix to SPie to SBud. This finding indicates that the highly skilled participants self-selected into compensation schemes offering high magnitudes of incentives. On average, as intended by the random assignment procedure, the participants in the assignment treatment (N = 83) show levels of skill that are as high as those of the participants in the self-selection group (N = 82). We can see that the magnitude of incentive has a particularly positive influence on the self-selection treatment groups (SFix, SPie and SBud; cf. also Figure 3). Surprisingly, AFix’s mean productivity (50.30) is greater than that of APie (49.00), whereas, as expected, the highest mean in the assignment group is found in the ABud treatment (56.22).

\textit{Insert Table 1 about here}

\textit{Insert Figure 3 about here}

We test the hypothesis by conducting a linear regression, as presented in Estimate 2 of Table 2. SKILL and MOI2 are significant. The covariate skill predicts productivity to a significant degree. According to the estimate, the participants solve 0.99 anagrams in Round 2 for each anagram solved in Round 1. MOI2’s B value represents the shift in the estimated productivity if a participant works under the budget-based scheme instead of the fixed pay scheme. The estimated number of anagrams solved increases by 7.44 if the participants work under the budget-based scheme instead of the fixed pay scheme. This increase is significant at the 0.05 level \((t = 2.13)\). The comparison of the fixed pay condition with the piece rate condition (MOI1) is not significant \((t = 0.06)\). Estimate 2 explains 67.4 percent of the variance in productivity. The hierarchical F test between the full (Estimate 2) and the skill-only model (Estimate 1) is statistically significant \((p = 0.00)\). Consequently, compared with the skill-only model, MOI2 provides a particularly significant contribution to our ability to predict productivity, even though R\textsuperscript{2} only increases by 0.027 (i.e., from 0.647 to 0.674). Thus, the results indicate that magnitude of incentive has a significant influence on productivity. However, only strong monetary incentives (i.e., the budget-based contract) exert an influence on productivity.

\textit{Insert Table 2 about here}

\textsuperscript{13} Refer to Figure 6 in the Appendix for histograms that show the participants’ skill distribution for each treatment group.
We expect freedom of choice to have a moderating effect in H1.B. As indicated by the differences in productivity change rates between the assignment and self-selection groups in Table 1 and Figure 3, freedom of choice appears to influence the relationship between magnitude of incentive and productivity. This finding indicates that a sorting effect may exist. In Estimate 2 of Table 2, the interaction terms and freedom of choice (FOC) are not significant. This result is due to the skill covariate. By conducting the regression without controlling for skill (not displayed), we find that the interaction components and FOC become significant at the 0.01 level ($R^2 = 0.254$). From these results, we can draw two corollaries. First, if individuals can self-select, then they sort according to their skill levels. Second, controlling for skill mainly controls for the sorting effects. Thus, if the sorting effects are not controlled for, then FOC highly influences productivity because the participants self-select based on their skill levels.

In H1.C, we predict that incentive and sorting effects change the variance in the number of anagrams solved. Whereas the standard deviation is 16.44 for the outcome in Round 1 (skill), this number increases to 22.03 (i.e., by 34 percent) for the outcome in Round 2 (productivity) (cf. Table 1). Consequently, working under incentives resulted in more widely distributed outcomes, as expected. If we account for sorting and compare the standard deviations of the productivity levels in the assignment groups (22.10) and the self-selection groups (22.09), then we find that the overall variation is almost equal. However, the patterns within the three magnitudes of incentive for each freedom of choice condition seem to be different. AFix has the highest treatment group variance, ABud has the lowest, and APie lies in between the two. Thus, within the treatment groups, the assignment groups have higher variance than the self-selection groups (SFix, SPie and SBud). Regardless, the total variance of the self-selection groups is the same as that of the assignment groups, as the self-selection groups vary around means that are relatively wide apart from each other. This finding shows that the participants’ sorting process yields groups that are relatively equal in productivity, whereas forcing incentives on the participants yields relatively heterogeneous productivity distributions. Thus, freedom of choice might have increased the participants’ commitment to their particular compensation schemes. Eriksson et al. (2009) highlight a related issue. They analyze effort variability in tournaments and find that it is

\[\text{...}\]

\[\text{\footnotesize\textsuperscript{14} The distributions are illustrated in Figure 7 in the Appendix, which depicts the frequencies of productivity levels during the 10-minute work period for the six different treatment groups.}\]
lower when agents can choose whether to work under a tournament format. This observation is explained by agents who exhibit unstable and extreme effort levels and who tend to stay out of the tournament.

Our research setting allows us to discriminate between incentive and sorting effects to a certain degree. To deepen this analysis, we shall refer to Figure 4. In both the fixed pay and budget-based conditions, the sorting effects are larger than the incentive effects, whereas in the piece rate condition, the incentive effects have a larger weight. The sorting effect is strongly negative in the fixed pay condition but strongly positive in the budget-based condition. Overall, the sum of the absolute values is larger for the sorting effects than for the incentive effects.15

5.2 Contingencies of Incentive Effects

As shown in Table 3, we conduct a multiple regression to determine the impacts of the magnitude of incentive, intrinsic motivation factors (i.e., need for achievement, need for cognition, interest, challenge and anxiety) and locus of control on productivity while controlling for skill in accordance with H2.A and H2.B. The model predicts productivity significantly better than the skill-only model in Estimate 1 of Table 2 (p = 0.00) and explains 70.3 percent of the variance in productivity. Magnitude of incentive only has an influence when strong incentives are in place (MOI2). Furthermore, challenge is highly significant (t = 2.94). When measured in standardized units β, the effect size of challenge is almost as high as that of MOI2 (β_{CHALLENGE} = 0.147, β_{MOI2} = 0.148). We find no productivity effects for the individuals differing in their needs for achievement and needs for cognition.16

---

15 The underlying data in Figure 4 need to be regarded carefully because they are derived from different subgroups and only compare the absolute percentages. An alternative rationale considering percentages might be formulated as follows. On average, productivity increases by 24 percent when controlling for skill under the budget-based scheme. Because of the sorting effects, the participants in SBud are 40 percent more skilled than the participants in ABud. In this case, the performance difference induced by the sorting effect is approximately 1.7 times higher than the performance difference induced by the incentive scheme (0.4 / 0.24).

16 The interactions between magnitude of incentive and need for achievement as well as need for cognition are not significant either and are not displayed. There is slight evidence for a crowding out effect in that the participants with higher interest in the task tend to have a higher performance improvement rate (i.e., skill / productivity) in the fixed conditions than in the piece rate or budget-based conditions. Spearman’s Rho test, which provides ratings on a 1-5 scale, was used to evaluate the following statement: “After having read the instructions, the anagram task seemed very interesting to me.” The performance improvement is highest for the AFix treatment group (Rho =

---
Anxiety and internal control conviction have a negative influence on productivity. We derived \(H2.B\) by rating the anagram task as a non-routine and complex task. This finding for internal control conviction is somewhat surprising. However, before the productivity session in Round 2 took place, the participants had already spent 13 minutes solving anagrams in Round 1 (i.e., 3 minutes in the practice session and 10 minutes in the skill session). Because the individual anagram tasks were relatively short and sometimes could be solved within seconds, the participants might have felt some sense of routine after having worked on the tasks for a while. Spector (1982) states that for tasks that are routine in nature and that demand strict adherence to rules and procedures, internals might show lower performances than externals. Hence, an effect contrary to our expectations might have occurred.\(^\text{17}\)

Overall, skill seems to have a paramount influence on productivity. Skill is the most important predictor in Estimate 2 in terms of standardized effect sizes and explains 64.8 percent of the variance in productivity (Estimate 1 of Table 2).

### 5.3 Contingencies of Sorting Effects

To analyze participants’ decisions regarding their incentive schemes (i.e., fixed compensation, piece rate compensation or budget-based compensation plans), we analyze the self-selection condition (i.e., the SFix, SPie and SBud groups). Table 4 presents the multinominal regression related to hypotheses \(H3.A\) to \(H3.D\). In the regression, we set the fixed pay condition as a reference group (coding: 1) and compare it with the piece rate pay condition (coding: 2) as well as the budget-based condition (coding: 3). As a result, we generate two equations termed ‘1 vs. 2’ and ‘1 vs. 3’. We calculate the comparison equation ‘2 vs. 3’ by subtracting the coefficient \(B_p\) of equation ‘1 vs. 2’ from the coefficient \(B_p\) of equation ‘1 vs. 3’: \(B_p\ (2 \text{ vs. } 3) = B_p\ (1 \text{ vs. } 3) - B_p\ (1 \text{ vs. } 2)\)
Nagelkerke’s pseudo $R^2 (= 0.681)$ of the full model (Estimate 2 of Table 4) is sizable and larger than the $R^2 (= 0.589)$ of the skill-only model in Estimate 1.

With respect to the decision between the fixed pay and the budget-based pay plans (1 vs. 3) in Estimate 2, a participant’s likelihood of choosing the budget-based pay scheme instead of the fixed pay scheme increases by 6.8 percent for each additional skill unit possessed by the participant.\textsuperscript{18} \textit{Ceteris paribus}, for each additional unit in need for achievement, internal control conviction and risk aversion, the probability increases by 12.9 percent, decreases by 8.1 percent and decreases by 13.8 percent, respectively.\textsuperscript{19} Thus, the predictions of $H3.A$ to $H3.D$ are supported.

### 5.4 Empirical Thresholds

In the following section, we investigate the empirical indifference thresholds $\pi^1$, $\pi^2$ and $\pi^3$ and the impacts of \textit{intrinsic motivation}, \textit{locus of control} and \textit{risk preference} on these thresholds based on the data of the logistic regressions performed in this study.

To specify the expectation, we use the contract parameters $\alpha, \beta, \gamma$ and $\delta$. As explained above, in the present experiment, $\alpha = 10$ Euros, $\beta = 0.23$ Euros, $\gamma = 28$ Euros, and $\delta = 4$ Euros. As a result, there are intersections between $W_F$ and $W_p$ at 43.48, between $W_F$ and $W_B$ at 56 and between $W_P$ and $W_B$ at 17.39 and 56 (cf. Figure 2). Because we measure productivity in terms of natural numbers, we can infer the following statements for the participants in the experimental situation, where they must choose among the three contract alternatives. When the value of productivity $x_i$ ranges from 0 to 43, the pay of $W_F$ is highest. From 44 to 55, the pay of $W_P$ is

\textsuperscript{18} The data in Table 1 and the histograms in Figure 6 in the Appendix already indicate that a positive relationship exists between skill and the magnitude of incentive in the selected contract because the mean skill levels increase from SFix (34.24) to SPie (49.09) to SBud (61.43). This increase is assumed to stem from the conscious decisions of higher-skilled individuals to choose compensation schemes with higher magnitudes of incentive.

\textsuperscript{19} Because the $B$ values and marginal effects are not standardized, we need to consider them in relation to the variation and range of the measures to judge their relative importance. The standard deviations are as follows: $\text{SKILL} = 16.44$, $\text{NFA} = 1.99$, $\text{NFC} = 10.88$, $\text{ICP} = 4.14$, $\text{RISKAV} = 1.48$. The need for cognition is not significant in the full model but predicts the decision to a significant degree in the model: magnitude of incentive = $f$ (skill, need for cognition) in the 1 vs. 2 comparison at the 0.1 level. This result is ostensibly due to the correlation effects. Guided by the research design, where the consideration of threshold values plays an important role, and by the nature of logistic functions, which can best be approximated by linear functions around zero logits ($z$), we generally evaluate the coefficients’ effect sizes by calculating the marginal effects based on zero logits.
highest. From 56 onwards, the pay of $W_B$ is highest. Consequently, the relevant crossing points are the intersections between $W_F$ and $W_P$ at 43.49 and between $W_F$ and $W_B$ at 56. Assuming that individuals base their decisions solely on the aim of maximizing their pay and that they use their skill levels as their reference points, we can expect the participants to be indifferent with regard to the various payment schemes at the respective crossing points. If a participant deletes $W_P$ from the decision set in the first step and decides between $W_F$ and $W_B$ solely based on the output in Round 1, then he/she theoretically chooses $W_B$ at and above a skill level of 56. This comparison might be fruitful, particularly because we cannot rule out that the participants made their final decisions between the fixed and budget-based compensation scheme after having decided against the piece rate scheme in the first step. This consideration seems particularly relevant given that the range between the lower (44) and upper thresholds (56) is not large.

When we set the logit $z_i$ to 0 in $z_i = -6.084 + 0.134 \text{SKILL}_i$ (cf. 1 vs. 2 comparison, Estimate 1, Table 4) and solve the resulting equation for $\text{SKILL}_i$, we find that $\text{SKILL}_i$ or $\pi^1 = 45.40$. Hence, at a skill level of 45.40, an individual’s estimated probability of selecting the piece rate contract instead of the fixed contract is 0.5. Consequently, the threshold ($\pi^1$) at which an individual is indecisive between the two payment alternatives is close to the theoretically expected threshold of 43.47. For the 2 vs. 3 comparison, $\pi^2$ is 55.88. Again, this value is close to the theoretically derived threshold value of 56. The threshold values that we estimated based on our empirical observations and their closeness to the theoretically expected threshold values support the validity of the indifference model developed above. Assuming that there is no piece rate contract and comparing the fixed pay contract with the budget-based contract (1 vs. 3) lead to an estimated $\pi^3$ of 50.00, which indicates the level at which the participants are indifferent between the budget-based and fixed compensation schemes.

Because the fixed pay contract and the budget-based contract are at opposing ends for the magnitude of incentive variable, the comparison (1 vs. 3) results in the highest significance values (Table 4). Figure 5 shows how the threshold of the 1 vs. 3 decision depends on the participants’ internal control preferences, risk aversion and needs for achievement (i.e., intrinsic motivation). The graphs show that the higher the level of preference for the internal control and

---

20 We assume that $0 < x_i < 120$. One participant in the SBud condition even solved more than 120 anagrams.
risk aversion are, the higher the threshold at which the individuals are indifferent between the fixed and the budget-based contract and, thus, the lower the likelihood of choosing the budget-based contract. In contrast, the higher the need for achievement, the lower the respective threshold is. Thus, the empirical observation supports the model’s prediction that the participants with equal skill levels are influenced by their respective individual preferences when making their decisions. Internally oriented and risk-averse individuals select the fixed pay plan rather than the budget-based compensation plan, whereas high-achievers select the budget-based scheme.

\textit{Insert Figure 5 about here}

6 Discussion

In sum, certain factors impact either the incentive or the sorting effects. We find that risk aversion is a determinant of self-selection behavior, whereas anxiety is an influencing factor on the incentive effect.\textsuperscript{21} Additionally, three factors are involved in both the incentive and sorting effects. First, we need to emphasize that the most important factor in both effects is skill. It strongly determines the participants’ choice of contracts and is highly correlated with productivity. Second, the magnitude of incentive factor plays a twofold role. It is an important determinant of productivity in the incentive effect and needs to exist for a sorting effect to occur. However, the magnitude of incentive is the result and not the determinant of the self-selection process. Third, we found that locus of control influences both the sorting and incentive effects. We find that the decisions made by the participants with equal skill levels are influenced by their preferences regarding their locus of control. The internals select the fixed pay plan rather than the budget-based compensation scheme. Given that internals are represented by the fixed pay group rather than by the budget-based compensation group, internals might show lower than average productivity levels because the fixed pay scheme has a lower magnitude of incentive. Interestingly, we find that the internal control conviction also leads to lower productivity when we control for the magnitude of incentive.

\textsuperscript{21} In terms of an affective state, anxiety may be a determinant of selection behavior as well. However, because we measure anxiety after the contract selection takes place, we cannot consider anxiety to be a determinant of the contract selection process. Thus, the tested influencing factors are not a comprehensive list of the variables of the incentive and sorting effects but rather a result of the research design, the identified research gaps and the hypotheses in this study.
While considering the two previously stated objectives, we discuss the results of our study in the following section. We discuss labor economics and the organizational implications of our study. With respect to *Objective 1*, the results confirm the presence of incentive effects driven by the magnitude of incentive in the selected contract. The participants are more productive under the budget-based compensation scheme than under the fixed scheme. Interestingly, we do not find universal evidence that the piece rate scheme, which ranges in the middle of the spectrum in terms of magnitude of incentive, is significantly better at motivating higher productivity than the fixed pay scheme. The data indicate that skill-based sorting tends to have a greater effect on the final output than the direct incentive effects. In comparison, Lazear (2000) states that the incentive and sorting effects are equally responsible for the total productivity increase of 44 percent found in his study. Both findings indicate that the sorting effects induced by incentives can explain a substantial part of productivity. This finding implies that sorting effects, which often seem to be ignored, need more attention (cf. Chiappori and Salanić 2003; Eriksson and Villeval 2008; Lazear 2004). With regard to management practices, we recommend that managers externally promote a well-designed incentive system to create sorting effects and to attract highly skilled employees. Doing so might help to not only overcome negative fluctuations due to employees who disapprove of the incentive system leaving the company but also boost overall business productivity.

Additionally, we find that working under monetary incentives results in more widely distributed outcomes than working without monetary incentives. Furthermore, the productivity variance within the self-selected groups is smaller than the variance within the assigned groups. Eriksson et al. (2009) highlight a related issue. They analyze the effort variability in tournaments and find that it is lower when the agents can choose whether to work under a tournament format. This issue implies that when incentives are forced upon the workforce, these incentives might have divergent productivity effects in the short run. Greater productivity differences lead to greater wage gaps when the variable pay is in place. These gaps can lead to struggles within a company’s workforce (cf. Clark et al. 2010). However, as the workforce becomes sorted over time, the differences in the levels of output can once again become balanced, and the workforce may become more homogeneous.
Objective 2 regards the incentive effects separately from the sorting effects. Apart from magnitude of incentive and skill, we identified other factors as the contingencies for incentive effects. The influence of intrinsic motivation and the weak effects of piece rate pay underline the notion that merely increasing magnitudes of incentive does not universally motivate employees to achieve higher productivity levels. In line with Gneezy and Rustichini (2000), we find that if monetary incentives are used, then their magnitude should be high enough to yield productivity effects. The negative effect of internal control conviction on productivity underlines the importance of a strong employee-task fit to the efficient allocation of employees. For example, the direct effects that preferences regarding locus of control have on productivity depend on the task itself. Thus, both the firm and the employee benefit if individual preferences are aligned with the available tasks in a company.

As to the contingencies of sorting effects, the results show that sorting effects depend not only on the relationship between skill and magnitude of incentive but also on other factors. This finding has been corroborated by the impact that risk allocation effects have on contracting. That is, less risk-averse individuals tend to select performance-based compensation plans rather than fixed compensation plans.

The sorting effect is also interesting from another perspective (i.e., when changing the level of analysis from the individual to the organization as the location of human activity). Schneider (1987) claims that sorting effects even affect the culture of an organization in the long run because people are the most important determinant of organizational behavior. Consider that if skill is equal, then the higher the magnitude of incentive in the compensation schemes, the lower the risk aversion of the individuals who select these schemes will be. This finding implies that in the long run, organizations that offer risky contracts to their employees will have a workforce that is relatively risk-seeking not only in terms of their own compensation contracts but also in terms of the decisions that they make on behalf of their organization. Thus, because people are an important determinant of organizations, the people’s characteristics shape the way organizations act in their business environments. In the end, incentives not only influence an organization’s ability to attract employees but also the characteristics of the organization itself.

This study shows that there is a connection between variable pay and risk preference such that variable pay attracts employees who engage in risk-seeking behavior. Consequently, the study
supports some researchers’ arguments that the financial industry’s payment practices have contributed to the development of the financial crises that started in 2008 (cf. Diamond and Rajan 2009). Specifically, sorting might have contributed to a composition of employees in the financial industry that is characterized by less-than-average risk aversion, which resulted in increased risk-taking behavior.

The results also imply that individuals with higher needs for achievement tend to select contracts with higher magnitudes of incentive. Given the aforementioned sorting effects, an organization that attracts individuals with higher needs for achievement might become more ambitious. Because previous scholars have shown that the level of need for achievement and the tightness of goal levels are positively related (cf. Atkinson 1957; Lee et al. 1997), organizations might tend to set more ambitious goals and strive for superior performance if they attract high-need-for-achievement individuals.

We come across another controversy when we consider the previous findings related to the need for achievement and the crowding out effect. Because the underlying results imply that higher needs for achievement leads to the selection of higher magnitudes of incentive and that, in turn, higher magnitudes of incentive lead to higher productivity levels, we need to put Vecchio’s (1982) finding and the crowding out effect into perspective. High-achievement individuals (i.e., individuals who possess high levels of intrinsic motivation) may not be as strongly motivated by monetary incentives as low-achievers. However, if high-achievement individuals can select their contracts, then they opt for contracts with higher monetary incentives. Thus, this finding somewhat differs from Delfgaauw and Dur’s (2007) analytic results. However, in line with their reasoning, we think that the nature of the effects strongly depends on the operationalization of intrinsic motivation. Thus, it is important to note that our results and rationale are based on our interpretation of need for achievement as a construct of intrinsic motivation.

Additionally, we found that an individual’s tendency to believe that a situation is under his or her own control influences the sorting effects. The higher this tendency (i.e., the more an individual is internally controlled), the higher the likelihood is that he/she selects a contract with fewer incentives. This effect is explained by an internal’s preference for organizational independence. Because incentives tie an employee’s actions to a superior’s goal, they limit the organizational independence of the employee. Thus, we can view an employee’s freedom of actions as being
bound by his or her incentives. Because an employee is bound with respect to the budget in a budget-based contract or by piece rates but is unbound in a fixed pay contract, an employee who tends to have a strong degree of internal control conviction tends to prefer fixed pay contracts. In this case, employees might demand what we call a bondage premium.

If this sorting effect has an impact on the overall organizational culture in the long run (cf. Schneider 1987), then the locus of control dimension might determine whether organizations perceive their environments as under their control or determined by external market forces. Given that individuals’ characteristics spill over to organizational characteristics, the locus of control belief might influence the extent to which organizations emphasize decision facilitating systems, such as decision support systems or business intelligence systems, in their management accounting practices. These systems may be able to provide internal and external information upon which managers may base their decisions. The belief regarding the extent of influence that these decisions have within the environment might influence the amount of effort that managers are willing to exert in building decision facilitating systems.

Because the results show that individuals possessing different attributes are attracted to different incentives, managers should take into account what kinds of individuals are attracted to which incentives to shape characteristics of the workforce and the whole organization, eventually. For better effectiveness of incentive systems, incentive and sorting effects need to be regarded.

7 Limitations and Outlook

The results and the setting of the experiment are subject to several limitations and offer ideas for a variety of extensions or considerations regarding certain issues.

The monetary attributes of the contracts involved in the experiment have influenced risk-taking behavior. However, in reality, a wide variety of factors might influence employees’ perceptions of the risks involved in a contract. For example, employment protection legislation might be another important risk factor for employees seeking a new contract. It might be interesting to observe whether the decision to choose a certain contract is also influenced by the interaction between employment protection legislation and risk preferences. This topic may become more relevant in light of the increasingly globalized nature of the job market. Additionally, future researchers need to further consider the measurement of risk preferences because of the great
number and variation of those measures. With respect to risk perception, future studies aiming to investigate the direct influence of risk perception on incentive choice might consider allowing their participants to independently assess the risk levels of each contract before the researchers perform treatments on their participants.

We also raise the question of causality with particular respect to the locus of control. This study has analyzed the effects of locus of control on individuals’ behavior while assuming that locus of control is a stable trait. However, in the long run, another chain of thought might be valid as well. In line with Burks et al. (2009), who explain that different work environments can prompt different norms that affect behavior, the characteristics of a company might shape individuals’ attributes. For instance, an employee’s preference regarding his or her locus of control might be affected by several company attributes. It might be interesting to observe whether the employees of bigger companies tend to have different preferences regarding their loci of control than the employees of smaller companies because of the differences in market power or market share between the two types of companies. Of course, sorting effects might play a role in this case as well. Future questions that might be raised include the following: Are public sector employees different from private sector employees in terms of locus of control? Are university professors different from chief executive officers in terms of locus of control? How do differences develop with respect to the length with which a person remains affiliated with a particular work environment?

To broaden the view, we put the output-based approach of this study into perspective. Firms choose their compensation schemes by comparing not only the benefits but also the costs of each scheme to create cost-efficient incentive schemes. We observed the benefits on the productivity levels. We ignored the costs, which may consist of monetary payouts, measurement difficulties and undesirable risk allocation or quality problems. In our study, the average monetary payout of each compensation plan can serve as a cost measure. On average, the fixed-paid participants received 10 Euros (N = 65), the piece rate-paid individuals received 12.06 Euros (N = 52, SD = 4.26) and the budget-based-paid participants earned a mean of 20.50 Euros (N = 48, SD = 11.24). Thus, the budget-based-paid participants earned more than double the compensation of the fixed-paid individuals. Incorporating the costs of incentive schemes adds additional complexity to the research and might be another approach for future researchers.
Acknowledgements

We gratefully acknowledge the comments and suggestions of Robert J. Bloomfield, Marc W. Nelson and the ESA Conference Chicago 2011 participants.

Reference List


## Tables

**Table 1**: Skill and productivity per treatment variable and treatment group

<table>
<thead>
<tr>
<th>Treatment variables</th>
<th>Assignment, N = 83</th>
<th>Self-selection, N = 82</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freedom of choice</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SKILL = 44.49 (17.21)</td>
<td>SKILL = 45.37 (15.72)</td>
<td></td>
</tr>
<tr>
<td>PRY = 51.77 (22.10)</td>
<td>PRY = 52.22 (22.09)</td>
<td></td>
</tr>
<tr>
<td>Magnitude of incentive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>API = 1.16</td>
<td>API = 1.15</td>
<td></td>
</tr>
<tr>
<td>ASD = 0.99</td>
<td>ASD = 1.01</td>
<td></td>
</tr>
<tr>
<td>Fixed pay, N = 65</td>
<td>AFix, N = 27</td>
<td>SFix, N = 38</td>
</tr>
<tr>
<td>SKILL = 38.92 (15.07)</td>
<td>SKILL = 45.52 (17.28)</td>
<td>SKILL = 34.24 (11.34)</td>
</tr>
<tr>
<td>PRY = 42.65 (22.19)</td>
<td>PRY = 50.30 (26.96)</td>
<td>PRY = 37.21 (16.37)</td>
</tr>
<tr>
<td>API = 0.10</td>
<td>API = 0.11</td>
<td>API = 0.09</td>
</tr>
<tr>
<td>ASD = -0.13</td>
<td>ASD = 0.01</td>
<td>ASD = -0.24</td>
</tr>
<tr>
<td>Piece rate pay, N = 52</td>
<td>APie, N = 29</td>
<td>SPie, N = 23</td>
</tr>
<tr>
<td>SKILL = 46.25 (14.31)</td>
<td>SKILL = 44.00 (17.31)</td>
<td>SKILL = 49.09 (8.85)</td>
</tr>
<tr>
<td>PRY = 52.42 (18.46)</td>
<td>PRY = 49.00 (20.83)</td>
<td>PRY = 56.74 (14.25)</td>
</tr>
<tr>
<td>API = 0.13</td>
<td>API = 0.11</td>
<td>API = 0.16</td>
</tr>
<tr>
<td>ASD = 0.03</td>
<td>ASD = -0.02</td>
<td>ASD = 0.09</td>
</tr>
<tr>
<td>Budget-based pay, N = 48</td>
<td>ABud, N = 27</td>
<td>SBud, N = 21</td>
</tr>
<tr>
<td>SKILL = 51.63 (17.72)</td>
<td>SKILL = 44.00 (17.63)</td>
<td>SKILL = 61.43 (12.42)</td>
</tr>
<tr>
<td>PRY = 64.19 (19.50)</td>
<td>PRY = 56.22 (17.75)</td>
<td>PRY = 74.43 (16.98)</td>
</tr>
<tr>
<td>API = 0.24</td>
<td>API = 0.28</td>
<td>API = 0.21</td>
</tr>
<tr>
<td>ASD = 0.15</td>
<td>ASD = -0.02</td>
<td>ASD = 0.37</td>
</tr>
</tbody>
</table>

Overall: N = 165, SKILL = 44.93 (16.44), PRY = 51.99 (22.03), API = 1.16

Notes: Means and standard deviations (in parentheses) of selected variables for each experimental treatment group as well as experimental dimension. SKILL, number of anagrams solved correctly within ten minutes in Round 1; PRY, number of anagrams solved correctly within ten minutes in Round 2 (productivity); API, average performance increase (PRY / SKILL – 1); ASD, average skill difference (particular skill mean / sample skill mean for the 165 participants – 1); Even though the budget-based paid participants needed to solve only 56 anagrams correctly in order to receive the bonus, their mean productivity lies substantially above that level (64.19). Since the participants had no possibility to track their performance levels during the work task, it is likely that they worked as hard as possible without knowing when the target level was achieved. Responses to Question P.1 support this rationale. Being asked to describe their effort development in the 10 minutes work period, only eight participants
indicated that they erratically decreased their effort. Erratically decreasing the effort after having solved 56 anagrams would minimize effort and maximize pay. From these eight participants only two worked under the budget-based compensation scheme. Both were in the treatment SBud (one participant solved 70; the other 100 anagrams). Thus, there is no evidence for a systematic productivity bias in the budget-based scheme.

Table 2: Monetary incentives, freedom of choice and productivity

<table>
<thead>
<tr>
<th>Estimate</th>
<th>$R^2$</th>
<th>Coefficient name</th>
<th>$B$</th>
<th>SE</th>
<th>$\beta$</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.647***</td>
<td>(Constant)</td>
<td>3.548</td>
<td>2.981</td>
<td>1.190</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SKILL</td>
<td>1.078</td>
<td>0.062</td>
<td>0.805***</td>
<td>17.302</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Constant)</td>
<td>5.039</td>
<td>4.020</td>
<td>1.254</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SKILL</td>
<td>0.994</td>
<td>0.070</td>
<td>0.805***</td>
<td>14.260</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MOI1</td>
<td>0.214</td>
<td>3.430</td>
<td>0.005</td>
<td>0.062</td>
</tr>
<tr>
<td>2</td>
<td>0.674***</td>
<td>MOI2</td>
<td>7.436</td>
<td>3.490</td>
<td>0.154**</td>
<td>2.130</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FOC</td>
<td>-1.869</td>
<td>3.321</td>
<td>-0.043</td>
<td>-0.563</td>
</tr>
<tr>
<td></td>
<td></td>
<td>INTMOIFOC1</td>
<td>4.550</td>
<td>4.952</td>
<td>0.072</td>
<td>0.919</td>
</tr>
<tr>
<td></td>
<td></td>
<td>INTMOIFOC2</td>
<td>2.747</td>
<td>5.322</td>
<td>0.042</td>
<td>0.516</td>
</tr>
</tbody>
</table>

Notes: N = 165; dependent variable: Productivity, anagrams solved within 10 minutes in Round 2; independent variables: SKILL, anagrams solved within 10 minutes in Round 1; MOI1, magnitude of incentive dummy variable 1, with 0 = fixed pay condition and 1 = piece rate condition; MOI2, magnitude of incentive dummy variable 2, with 0 = fixed pay condition and 1 = budget-based condition; FOC, freedom of choice dummy variable, with 0 = assignment condition and 1 = self-selection condition; INTMOIFOC1 = FOC \cdot MOI1, interaction of FOC and MOI1; INTMOIFOC2 = FOC \cdot MOI2, interaction of FOC and MOI2; *, **, *** denote significance at the 10 percent, 5 percent and 1 percent levels, respectively (two-tailed tests); $R^2 = 0.647$ for Estimate 1 ($p = 0.00$), $\Delta R^2 = 0.027$ for Estimate 2 compared to Estimate 1 ($p = 0.00$); Max(VIF) = 3.16; Max(Cook’s Distance) = 0.47

Table 3: Individual contingency variables and productivity

<table>
<thead>
<tr>
<th>$R^2$</th>
<th>Coefficient name</th>
<th>$B$</th>
<th>SE</th>
<th>$\beta$</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.703***</td>
<td>(Constant)</td>
<td>17.741</td>
<td>10.316</td>
<td>1.720</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SKILL</td>
<td>0.959</td>
<td>0.066</td>
<td>0.721***</td>
<td>14.585</td>
</tr>
<tr>
<td></td>
<td>MOI1</td>
<td>2.749</td>
<td>2.398</td>
<td>0.058</td>
<td>1.147</td>
</tr>
<tr>
<td></td>
<td>MOI2</td>
<td>7.058</td>
<td>2.611</td>
<td>0.148***</td>
<td>2.703</td>
</tr>
<tr>
<td></td>
<td>NFA</td>
<td>0.093</td>
<td>0.560</td>
<td>0.009</td>
<td>0.167</td>
</tr>
<tr>
<td></td>
<td>NFC</td>
<td>-0.097</td>
<td>0.110</td>
<td>-0.046</td>
<td>-0.880</td>
</tr>
<tr>
<td></td>
<td>CHALLENGE</td>
<td>1.012</td>
<td>0.345</td>
<td>0.147***</td>
<td>2.935</td>
</tr>
<tr>
<td></td>
<td>ANXIETY</td>
<td>-1.485</td>
<td>0.546</td>
<td>-0.130***</td>
<td>-2.720</td>
</tr>
<tr>
<td></td>
<td>INTEREST</td>
<td>0.131</td>
<td>0.250</td>
<td>0.024</td>
<td>0.524</td>
</tr>
<tr>
<td></td>
<td>ICP</td>
<td>-0.577</td>
<td>0.267</td>
<td>-0.105**</td>
<td>-2.167</td>
</tr>
</tbody>
</table>

Notes: N = 164; dependent variable: Productivity, anagrams solved within 10 minutes in Round 2; independent variables: SKILL, anagrams solved within 10 minutes in Round 1; MOI1, magnitude of incentive dummy variable 1, with 0 = fixed pay condition and 1 = piece rate condition; MOI2, magnitude of incentive dummy variable 2, with 0 = fixed pay condition and 1 = budget-based condition; NFA, need for achievement level; NFC, need for
cognition level; CHALLENGE, level of challenge; ANXIETY, level of anxiety; INTEREST, level of interest; ICP, tendency to believe that outcomes are consequences of one’s own actions; *, **, *** denote significance at the 10 percent, 5 percent and 1 percent levels, respectively (two-tailed tests); Max(VIF) = 1.54; Max(Cook’s Distance) = 0.79

**Table 4: Individual contingency variables and decisions for incentives**

<table>
<thead>
<tr>
<th>Comparison</th>
<th>1 vs. 2</th>
<th>1 vs. 3</th>
<th>2 vs. 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Estimate 1</strong></td>
<td>Intercept</td>
<td>SKILL</td>
<td>Intercept</td>
</tr>
<tr>
<td><strong>B</strong></td>
<td>-6.084***</td>
<td>0.134***</td>
<td>-11.951***</td>
</tr>
<tr>
<td>SE</td>
<td>1.569</td>
<td>0.036</td>
<td>2.425</td>
</tr>
<tr>
<td>Marginal effect</td>
<td>0.034</td>
<td>0.049</td>
<td>0.059</td>
</tr>
</tbody>
</table>

Notes: dependent variable: MAGOI, magnitude of incentive, MAGOI = 1, fixed pay (N = 38), MAGOI = 2, piece rate pay (N = 23), or MAGOI = 3, budget-based pay (N = 21); independent variable: SKILL, anagrams solved within 10 minutes in Round 1; NFC, need for cognition level; NFA, need for achievement level; RISKAV, risk aversion level; ICP, internal locus of control preference level; *, **, *** denote significance at the 10 percent, 5 percent and 1 percent levels, respectively (two-tailed tests); Nagelkerke’s pseudo $R^2$ for Estimate 1 = 0.589, Estimate 2 = 0.681;
Figures

Figure 1: Overview of the experimental proceedings

![Diagram of experimental proceedings]

Figure 2: Compensation schemes

![Graph showing compensation schemes]

Notes: Payment in Euros is displayed as a function of productivity in anagrams solved for the three magnitudes of incentive conditions (fixed pay, piece rate pay or budget-based pay). For illustration purposes, the budget-based pay function is drawn as continuous vertical line at the step, although only one y-value exists at the x-value of 56, i.e. 28.
**Figure 3**: Productivity in anagrams per treatment group

Notes: The bar chart plots mean productivity in anagrams (y-axis) across the participants per treatment condition (x-axis). The black error lines depict 95 percent confidence intervals. Assignment groups: fixed pay (AFix), piece rate pay (APie) and budget-based pay (ABud); self-selection groups: fixed pay (SFix), piece rate pay (SPie) and budget-based pay (SBud);

**Figure 4**: Comparison of sorting and incentive effects

Notes: The sorting effects reflect the average skill difference values (ASD) of SFix, SPie and SBud from Table 1. The incentive effects reflect the average performance increase values (API) of the fixed, piece rate and budget-based pay conditions including the assigned and free-to-choose participants from Table 1.
Figure 5: Influence of locus of control, risk and intrinsic motivation on the skill threshold ‘1 vs. 3’

Notes: The figure depicts the anagrams skill threshold on the y-axis and the individual attribute level based on the inventories ICP, RISKAV and NFA measured in standard deviations from the mean on the x-axis. The following means (SDs) apply: ICP = 35.90 (4.03), N = 165; RISKAV = 5.75 (1.55), N = 162; NFA = 7.73 (2.03), N = 162. The graphs are based on the 1 vs. 3 comparisons in the multinomial regression models: magnitude of incentive = \( f \) (intercept, skill, individual attribute); Explicitly: \( z_i = -4.874 - 1.502 \text{ICPi} + 0.235 \text{SKILL}_i \), \( z_i = -9.808 + 0.258 \text{SKILL}_i - 0.530 \text{RISKAV}_i \) and \( z_i = -15.750 + 0.253 \text{SKILL}_i + 0.420 \text{NFA}_i \). In the equations \( z_i \) is set to zero and solved for \( \text{SKILL}_i \). This leads to: \( \text{SKILL}_i = 20.740 + 0.800 \text{ICPi} \), \( \text{SKILL}_i = 38.015 + 2.054 \text{RISKAV}_i \) and \( \text{SKILL}_i = 62.253 - 1.660 \text{NFA}_i \). Inserting the individual preference levels into the equations results in the dashed graphs. NFC is not shown as it is not significant in Table 4. The continuous graph reflects the skill threshold of the 1 vs. 3 comparison independent of ICP, RISKAV and NFA resulting from \( z_i = -11.951 + 0.239 \text{SKILL}_i \), as shown in Estimate 1 of Table 4.

---

Notes: The figure depicts the anagrams skill threshold on the y-axis and the individual attribute level based on the inventories ICP, RISKAV and NFA measured in standard deviations from the mean on the x-axis. The following means (SDs) apply: ICP = 35.90 (4.03), N = 165; RISKAV = 5.75 (1.55), N = 162; NFA = 7.73 (2.03), N = 162. The graphs are based on the 1 vs. 3 comparisons in the multinomial regression models: magnitude of incentive = \( f \) (intercept, skill, individual attribute); Explicitly: \( z_i = -4.874 - 1.502 \text{ICPi} + 0.235 \text{SKILL}_i \), \( z_i = -9.808 + 0.258 \text{SKILL}_i - 0.530 \text{RISKAV}_i \) and \( z_i = -15.750 + 0.253 \text{SKILL}_i + 0.420 \text{NFA}_i \). In the equations \( z_i \) is set to zero and solved for \( \text{SKILL}_i \). This leads to: \( \text{SKILL}_i = 20.740 + 0.800 \text{ICPi} \), \( \text{SKILL}_i = 38.015 + 2.054 \text{RISKAV}_i \) and \( \text{SKILL}_i = 62.253 - 1.660 \text{NFA}_i \). Inserting the individual preference levels into the equations results in the dashed graphs. NFC is not shown as it is not significant in Table 4. The continuous graph reflects the skill threshold of the 1 vs. 3 comparison independent of ICP, RISKAV and NFA resulting from \( z_i = -11.951 + 0.239 \text{SKILL}_i \), as shown in Estimate 1 of Table 4.
## Appendix

**Table 5: Descriptive statistics of individual preference variables**

<table>
<thead>
<tr>
<th>Number of items in inventory</th>
<th>N</th>
<th>Min.</th>
<th>Max.</th>
<th>Mean</th>
<th>SD</th>
<th>Cronbach’s alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>SKILL</td>
<td>-</td>
<td>165</td>
<td>7</td>
<td>96</td>
<td>44.93</td>
<td>16.44</td>
</tr>
<tr>
<td>NFA</td>
<td>13</td>
<td>162</td>
<td>4</td>
<td>13</td>
<td>7.73</td>
<td>2.03</td>
</tr>
<tr>
<td>NFC</td>
<td>14</td>
<td>164</td>
<td>-29</td>
<td>36</td>
<td>14.62</td>
<td>10.36</td>
</tr>
<tr>
<td>CHALLENGE</td>
<td>4</td>
<td>165</td>
<td>4</td>
<td>20</td>
<td>15.52</td>
<td>3.19</td>
</tr>
<tr>
<td>ANXIETY</td>
<td>2</td>
<td>165</td>
<td>2</td>
<td>10</td>
<td>4.96</td>
<td>1.94</td>
</tr>
<tr>
<td>INTEREST</td>
<td>4</td>
<td>165</td>
<td>4</td>
<td>20</td>
<td>12.82</td>
<td>4.09</td>
</tr>
<tr>
<td>ICP</td>
<td>8</td>
<td>165</td>
<td>26</td>
<td>46</td>
<td>35.90</td>
<td>4.03</td>
</tr>
<tr>
<td>RISKAV</td>
<td>-</td>
<td>162</td>
<td>3</td>
<td>9</td>
<td>5.75</td>
<td>1.55</td>
</tr>
</tbody>
</table>

Notes: Participants who successfully completed both rounds are considered in these statistics (N = 165). SKILL, number of anagrams solved correctly within ten minutes in Round 1; NFA, need for achievement level; NFC, need for cognition level; CHALLENGE, level of challenge; ANXIETY, level of anxiety; INTEREST, level of interest; ICP, tendency to believe that outcomes are consequences of one’s own actions; RISKAV, risk aversion level. If a respondent does not answer an item, the corresponding scale for the particular participant is not taken into account. NFA’s Cronbach’s alpha is equivalent to the Kuder-Richardson 20 procedure, because of the binary response format. The binary response format might be a reason why the value is the lowest among the reliability indicators presented.
Figure 6: Histograms of participants’ skill per treatment group

Notes: N = 165; For each treatment group a histogram is depicted, which shows skill levels on the x-axis and frequency of occurrence in percent on the y-axis. The dashed vertical line denotes mean treatment group skill level. The continuous vertical line serves as reference line and is at the level of the target in the budget-based scheme at 56 anagrams. The corresponding mean and SD values are given in Table 1.
Figure 7: Histograms of participants’ productivity per treatment group

Notes: N = 165; For each treatment group, a histogram is depicted which shows productivity levels on the x-axis and frequency of occurrence in percent on the y-axis. The dashed vertical line denotes mean treatment group productivity level. The continuous vertical line serves as reference line and is at the level of the target in the budget-based scheme at 56 anagrams. The corresponding mean and SD values are given in Table 1.