TECHNISCHE UNIVERSITÄT MÜNCHEN

Fakultät für Wirtschaftswissenschaften

Lehrstuhl für Unternehmensführung

The Impact of Monetary and Non-Monetary Incentives: An Experimental Analysis

Hanna Maria Sittenthaler

Vollständiger Abdruck der von der Fakultät für Wirtschaftswissenschaften der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Wirtschaftswissenschaften (Dr. rer. pol.) genehmigten Dissertation.

Vorsitzende:	Univ.-Prof. Dr. Hanna Hottenrott
Prüfer der Dissertation:	1. Univ.-Prof. Dr. Alwine Mohnen
	2. Univ.-Prof. Dr. Gunther Friedl

Die Dissertation wurde am 27.06.2019 bei der Technischen Universität München eingereicht und durch die Fakultät für Wirtschaftswissenschaften am 15.09.2019 angenommen.

To my parents

Acknowledgements

Writing a dissertation is like going on a long journey. It is a very exciting, enjoyable, and interesting time, but also challenging and full of lessons I am grateful to have learned. I would like to take this opportunity to thank all those who have accompanied and supported me on this academic journey during the last years.

First of all, I would like to express my deepest gratitude to my scientific advisor *Prof. Dr. Alwine Mohnen* for guiding and supporting me throughout all phases of this dissertation with her experience and knowledge. I am particularly thankful for her giving me the freedom to pursue my own research ideas and focus on the topics I am interested in. Thank you for your honest feedback, continuous support and your trust in my work.

I further would like to thank *Prof. Dr. Gunther Friedl* for serving as my second advisor and *Prof. Dr. Hanna Hottenrott* for acting as chairperson of the examination board for this thesis.

Moreover, I would like to thank my former and current colleagues at the Chair of Corporate Management: *Dr. Sven Asmus, Jan-Niklas Bartel, Nadine Bernklau, Dr. Dominik Doll, Christoph Drobner, René Gassen, Saskia Hausler, Anja Holzgethan, Dr. Sebastian Ihrig, Maximilian Kern, Karina Konieczny, Laura Lang, Dr. Sabrina Lehner, Verena Mückenhausen, Dr. habil. Andreas Ostermaier, Stefan Pabst, Dr. Maximilian Rupprecht, Dr. Dominik Schall, Nadeschda Schönauer, Arne-Kristian Schulz, Nevena Toporova, Katharina Wagner*, and *Martina Wayand*. I am particularly thankful for their valuable support, helpful discussions, and encouraging conversations, but also for the enjoyable working atmosphere, lunch breaks, and get-togethers. My further gratitude goes to *Virginia Herbst* for her valuable input and support. A special thanks goes to *Laila El-Kady* and *Renate Rast* for their helpful administrative support during the last years. In addition, I would also like to thank the *student assistants of the Chair of Corporate Management* for supporting me.

Moreover, I would like to thank my *closest friends* for listening and encouraging me whenever needed.

My deepest and most sincere gratitude goes to my family. Without the unconditional support of *my parents*, *my sister Sandra*, and *my partner Carsten* this journey would not have been possible. Thank you for your love, your patience and your encouragement. You have always believed in me and given me strength and support. Thank you from the bottom of my heart for always being there for me.

Table of contents

List of figures

List of tables

List of abbreviations

CEO	Chief Executive Officer
COPE	Colloquium on Personnel Economics
DiD	Difference-in-Differences
GEABA	German Economic Association of Business Administration
HR	Human Resources
HRM	Human Resource Management
IAB	Institut für Arbeitsmarkt- und Berufsforschung
LEN	Linear Exponential Normal
OLS	Ordinary Least Squares
ORSEE	Online Recruitment System for Economic Experiments
ZEW	Zentrum für Europäische Wirtschaftsforschung

Abstract

This thesis investigates the impact of incentives from a behavioral economics perspective using an experimental approach. Firstly, it analyzes the impact of variable payment cuts on performance. Secondly, it investigates the effectiveness of monetary incentives, non-monetary incentives, and a combination of these two, and sheds light on the impact of gender differences. Thirdly, it analyzes the effectiveness of different non-monetary incentives and a cafeteria-style incentive system. Understanding the underlying psychological mechanisms of incentives is crucial to implement them effectively.

Kurzfassung (German abstract)

Die vorliegende Dissertation untersucht anhand von Experimenten die Auswirkungen von Anreizen aus verhaltensökonomischer Sicht. Erstens wird der Einfluss von Bonuskürzungen auf die Performance analysiert. Zweitens wird die Wirksamkeit von monetären Anreizen, nicht-monetären Anreizen und einer Kombination beider untersucht und der Einfluss von Genderunterschieden betrachtet. Drittens wird die Wirksamkeit von verschiedenen nicht-monetären Anreizen und einem Cafeteria-Anreizsystem untersucht. Das Verständnis der zugrundeliegenden psychologischen Mechanismen von Anreizen ist wesentlich für deren effektive Implementierung.

1 Introduction

1.1 The need for incentives

"*Incentives are the essence of economics.*"
(Prendergast, 1999, p. 7)

Both researchers and practitioners have been investigating the role of incentives for a long time. In his paper, which provides an overview of the field, Prendergast (1999) emphasizes the importance of incentives in economics; they aim to maximize employee performance and align the interests of employers and employees. In general, incentives can be classified as monetary incentives, such as piece rates, bonuses, or profit sharing (Prendergast, 1999), and non-monetary incentives (Condly, Clark, & Stolovitch, 2003). The latter can be further distinguished as non-monetary tangible incentives, such as subsidized holiday trips, or restaurant coupons, and non-monetary intangible incentives, such as recognition, public praise, and positive performance appraisals (Condly et al., 2003). Nowadays, incentives are widely used in the workplace (Garbers & Konradt, 2014; Hammermann & Mohnen, 2014a; Incentive Research Foundation, 2017; Kauflin, 2017; Staufenbiel Institut & Kienbaum, 2017). However, there are several questions that arise when using incentives. Do incentives always increase work performance? Which type of incentive is the most effective − monetary or non-monetary? What is the optimal incentive level? Research on incentives has addressed these complex issues and related research questions for decades. Already in the 18th century, Adam Smith observed the issue of incentives by analyzing contracts and the division of labor in agriculture between landlords and their tenants and recognized the conflicting interests at play (Laffont & Martimort, 2002). Barnard (1938) also notes that incentives significantly affect the subjects' motivation. He was the first who made an attempt to propose a general theory of incentives in management (Laffont & Martimort, 2002) and stated that "[…] *in all sorts of organizations the affording of adequate incentives becomes the most definitely emphasized task in their existence*" (Barnard, 1938, p. 139). In doing so, Barnard recognizes the existence of different kinds of material and non-material incentives (Barnard, 1938; Laffont & Martimort, 2002).

Incentive theory originates from the situation of delegating a task to an agent who has private

information. Two types of private information may exist: (1) the action of the agent (e.g., employee) cannot be observed by the principal (e.g., employer) (*moral hazard* or *hidden action*), or (2) the agent has private knowledge, which means that there is asymmetric information between the agent and the principal (*adverse selection* or *hidden knowledge*) (Laffont & Martimort, 2002). Both the principal and the agent aim to maximize their individual utility and profit. Misaligned interests between these parties and asymmetric information, can lead to conflicts of interest when the principal delegates specific tasks to the agent. This issue is known in the literature as the *principal-agent problem* or *agency theory* and has been theoretically investigated in a large number of studies (Grossmann & Hart, 1983; Holmstrom, 1979; Jensen & Meckling, 1976; Prendergast, 1999; Ross, 1973; Spremann, 1987). According to the simplest standard agency models, such as the LEN-Model[1] also known as the Spremann-Model (Kräkel, 2007; Spremann, 1987), compensation consists of a fixed and a variable payment. Standard economic theory predicts that the higher the variable payment, the higher the performance. In contrast, a fixed wage provides no incentive and only ensures that the agent accepts the contract. Moreover, the principal must pay a risk premium to the risk-averse agent when variable payments are based on imperfect performance measures. Hence, using monetary incentives may lead to inefficient risk-sharing. The principal faces a tradeoff between risk-sharing associated with the loss of efficiency and the need to provide incentives to enhance performance (Bonner & Sprinkle, 2002; Holmstrom, 1989; Kräkel, 2007; Spremann, 1987). Nevertheless, standard economic theory emphasizes the relevance of variable, performance-based payments as they incentivize employees and ensure that they act in the best interest of the employer, thus contributing to the overall performance of the company (Bonner & Sprinkle, 2002; Kräkel, 2007; Prendergast, 1999).

The principal-agent model is possibly the most widely used framework for addressing compensation and incentive issues. However, it is also necessary to empirically analyze how subjects perform in response to different incentives and to analyze the incentives' underlying mechanisms to identify those incentives that are most suitable for enhancing employee performance. In the last two decades, several studies have shown that subjects' behavior differs from the predictions of standard economic theory owing to social preferences,

[1] The Linear-Exponential-Normal (LEN) Model is based on three core assumptions: linear compensation scheme, exponential utility function of the risk averse agent, and normally distributed performance measure (with regard to the error term) (Kräkel, 2007; Spremann, 1987).

behavioral “anomalies”, or a poor incentive structure (Fehr & Falk, 2002; Kamenica, 2012).

The topic of incentives has received a lot of attention from practitioners, as well as researchers; there is a large body of literature that analyzes, both theoretically and empirically, the effectiveness of monetary as well as non-monetary incentives (Condly et al., 2003; Jenkins, Mitra, Gupta, & Shaw, 1998; Prendergast, 1999). However, there are still questions regarding the effectiveness of incentives that have not yet been sufficiently addressed. This dissertation aims to contribute to the literature on incentives by addressing the identified gaps regarding the impact of monetary and non-monetary incentives from a behavioral economics perspective.

In particular, this research addresses whether and to what extent variable payment cuts of different levels affect performance and satisfaction (*Chapter 2*). To date the effects of variable payment cuts have been scarcely investigated empirically (Bareket-Bojmel, Hochman, & Ariely, 2017; Kawaguchi & Ohtake, 2007; Pouliakas, 2010; Smith, 2015).

Moreover, this thesis contributes to the literature on the effectiveness of monetary and non-monetary incentives (Condly et al., 2003; Hammermann & Mohnen, 2014b; Jeffrey, 2009; Kube, Maréchal, & Puppe, 2012) by analyzing whether monetary incentives, non-monetary incentives, or a combination of the two, lead to the highest employee performance. The impact of gender differences on the effectiveness of incentives is also examined (*Chapter 3*).

The effectiveness and underlying psychological mechanisms of non-monetary incentives have been rarely or, at least, insufficiently analyzed (Hammermann & Mohnen, 2014b; Jeffrey, 2009; Jeffrey & Shaffer, 2007; Kube et al., 2012). This thesis investigates whether different non-monetary incentives of equivalent value generate the same performance level and whether a cafeteria-style incentive system, whereby subjects are offered the choice between different non-monetary incentives, is effective (*Chapter 4*).

The above issues will be addressed with an experimental approach. Before presenting the main parts of the research in *Chapters 2* through *4*, I provide an overview of the relevant incentive literature (C*hapter 1.2*) and outline the research questions for my dissertation in greater detail (*Chapter 1.3*). Furthermore, I present the methodological approach (*Chapter 1.4*), as well as the key findings and main contributions of my thesis (*Chapter 1.5*).

1.2 Theoretical background and literature overview

This section provides an overview of the relevant literature regarding monetary and non-monetary incentives by focusing on behavioral and experimental economics. Emphasis is placed on studies that analyze the effects of incentives on individual performance. First, I introduce the literature on the impact of monetary incentives on performance. Then, I address the literature on non-monetary incentives and describe several explanatory approaches that analyze the effectiveness and motivational properties of these incentives.

1.2.1 Monetary incentives

Monetary incentives are, as the term applies, financial incentives and are often recommended for increasing employee performance (Bonner & Sprinkle, 2002). Nowadays, many companies link pay with performance, and use individual performance as the predominant basis for pay (Fang & Gerhardt, 2012). The effects of performance-related monetary incentives have been widely investigated in the literature. According to standard theoretical predictions of principal-agent models, higher monetary incentives should induce higher performance (Prendergast, 1999; Spremann, 1987). However, empirical studies show different results regarding the effectiveness of performance-related monetary incentives (Bonner, Hastie, Sprinkle, & Young, 2000).

On the one hand, several studies show that performance-related monetary incentives significantly enhance employees' performance in a variety of settings (Bonner et al., 2000; Condly et al., 2003; DellaVigna & Pope, 2018; Jenkins et al., 1998; Lazear, 2000; Prendergast, 1999; Shaw & Gupta, 2015). For example, using a meta-analysis, Condly et al. (2003) point out that financial incentives, on average increase performance by 27%. Furthermore, several field studies indicate that payment schemes based on piece rates are superior to fixed wages. For instance, Lazear (2000) shows by means of a data set of an auto glass company that the average performance of workers significantly increases after a change from hourly wages to piece rate pay. This increase in productivity may be attributed to: (1) an incentive effect, as people are more motivated to increase their effort, and (2) a selection effect, as a piece rate compensation scheme attracts more productive workers (Lazear, 2000). The effectiveness of performance-related pay is confirmed by several field studies (Bandiera, Barankay, & Rasul, 2007; Knez & Simester, 2001; Paarsch & Shearer,

2000). Paarsch and Shearer (2000), for example, show that piece rate contracts at a tree-planting firm lead to a 23% increase in productivity compared to fixed wage contracts. Nevertheless, they note that this increase in quantity is in part at the expense of quality. Freeman and Kleiner (2005) point out that although piece rates increase productivity, a piece rate compensation scheme requires monitoring the work quality, thus generating extra costs for the company.

On the other hand, some studies show that monetary incentives do not necessarily lead to an increase in performance, and can even have a detrimental effect (Camerer, Babcock, Loewenstein, & Thaler, 1997; Gneezy & Rustichini, 2000; Pokorny, 2008). An often mentioned reason for the negative impact of monetary incentives is *motivation crowding out* (Deci & Ryan, 1985; Deci & Ryan, 2002; Frey, 1997; Frey & Jegen, 2001; Gneezy & Rustichini, 2000; Mellström & Johannesson, 2008; Titmuss, 1970). According to Frey and Jegen (2001) the crowding out effect, also known as *corruption effec*t (Deci, 1975) or *cognitive evaluation theory* in psychology (Deci, Koestner, & Ryan, 1999; Deci & Ryan, 1985), is one of the most striking anomalies in economics. It suggests that monetary incentives may undermine intrinsic motivation and, therefore, decrease rather than increase performance. Deci (1972, p. 113) argues that "*a person is intrinsically motivated if he performs an activity for no apparent reward except the activity itself*". Titmuss (1970) states that monetary incentives crowd out the intrinsic motivation to donate blood. Mellström and Johannesson (2008) test this claim empirically and find evidence that, while for men the crowding-out effect is not statistically significant, women are less willing to donate blood when payment is introduced. Furthermore, Gneezy and Rustichini (2000) show that the relationship between monetary incentives and performance is not monotonic. Data from their field studies indicate a U-shaped relationship between incentive intensity and performance level: those who are offered small monetary incentives performed worse than those not offered incentives. In contrast, those who are offered large monetary incentives, exhibit higher performance. Nevertheless, Fang and Gerhardt (2012) point out the lack of convincing evidence on the occurrence of the crowding out effect in the workplace. Cameron, Banko, and Pierce (2001) use a meta-analysis to show that rewards do not necessarily have a pervasive negative effect on intrinsic motivation. This aspect is also underlined by DellaVigna and Pope (2018) who conduct an experiment on Amazon Mechanical Turk and show that low piece rates do not crowd out motivation. Moreover, in

contrast to Gneezy and Rustichini (2000), Pokorny (2008) finds an inverse U-shaped relationship between incentive intensity and performance level. These results are explained by using the theory of *reference-dependent preferences*, which can be traced back to Kahneman and Tversky (1979) who analyze decisions under uncertainty. Kahneman and Tversky (1979) show that subjects have an S-shaped value function, where the intersection is the reference point: below this point, the value function is convex (losses), while above it the slope is concave (gains). The key assumption is that the utility function does not only depend on the absolute payoff, but also on the relation between the payoff and an individual reference point. Payoffs below this reference point are perceived as losses, while payoffs above it are perceived as gains. Furthermore, the slope of the utility function flattens after the reference point is reached. Pokorny (2008) indicates that, as long as subjects are below their reference income, incentives increase their effort level, while beyond the reference income, higher incentives reduce subjects' performance. Besides Pokorny (2008), several other studies show that subjects' performance decreases after exceeding their reference income. This phenomenon indicates the relevance of reference-dependent preferences when analyzing the relationship between incentives and performance (Camerer et al., 1997; Chou, 2002; Fehr & Goette, 2007; Goette, Huffman, & Fehr, 2004).

Gneezy, Meier, and Rey-Biel (2011) point out that if the monetary incentive is large enough, then the relationship between monetary incentives and performance is positive, as the price effect dominates the crowding-out effect. However, monetary incentives should not be too high, as empirical evidence suggests that very high incentives can have a detrimental effect on performance − a phenomenon known as *choking under pressure* in psychology (Ariely, Gneezy, Loewenstein, & Mazar, 2009; Baumeister, 1984; Gneezy et al., 2011).

To conclude, performance-related monetary incentives are widely considered a suitable means to enhance employee performance. Nevertheless, contrary to the predictions of standard theory, the relationship between incentive intensity and effort levels is not necessarily monotonic due to *crowding out* or *reference-dependent preferences*.

Besides the research on performance-based monetary incentives, several studies focus on the incentive effect of fixed wages. According to standard economic assumptions, a fixed wage, unrelated to performance, provides no incentive (Prendergast, 1999; Spremann, 1987). However, several empirical studies challenge this assumption and analyze the impact of

exclusively paying fixed wages on performance. Most empirical evidence is based on the so-called gift exchange game[2], which goes back to Fehr et al. (1993) and depicts the relationship between different wage levels and effort. In contrast with the predictions of standard economic theory, empirical evidence shows that employers offer generous wages and workers on average respond to higher wage offers by choosing higher effort levels (Charness, 2004; Charness & Kuhn, 2007; Fehr et al., 1993; Fehr & Falk, 2002; Fehr, Gächter, & Kirchsteiger, 1997; Fehr, Kirchler, Weichbold, & Gächter, 1998). The concept of *reciprocity* (Charness & Rabin, 2002; Dufwenberg & Kirchsteiger, 2000; Dufwenberg & Kirchsteiger, 2004; Falk & Fischbacher, 2006; Fehr & Gächter, 2000; Rabin, 1993) and *inequity aversion* or *fairness* (Akerlof, 1982; Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999) are possible explanations for these results. *Reciprocity*, which is a basic principle of human behavior, means that friendly behavior is rewarded with a friendly action (*positive reciprocity*), whereas unfriendly behavior is punished with an unfriendly action (*negative reciprocity*) (Fehr & Falk, 2002; Fehr & Gächter, 2000). According to the gift exchange game, workers respond to a higher wage, which is regarded as a friendly action, with a higher effort level. Employers anticipate this behavior and offer higher wages to elicit these reciprocal responses. This kind of reciprocity-driven voluntary cooperation is an essential element in employment relationships (Fehr & Falk, 2002; Kube et al., 2012; Pereira, Silva, & Silva, 2006). The concept of *inequity aversion* assumes that the worker's utility depends not only on his or her payoff, but also on the payoff of the contracting party. Although individuals are largely driven by their payoff, they also wish to minimize inequality between the parties (Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999). According to the equity theory proposed by Adams (1963), a certain input-to-output ratio (e.g., the performance level to wage), in comparison to others, has to be preserved to obtain equity and ensure that individuals perceive a given wage as fair. More specifically, if an employee exerts greater effort than others, the employer should pay him or her a higher wage. The feeling of being treated fairly by the company is a crucial factor in the worker's effort decision. Following

[2] A standard *gift exchange game* is a sequential game played between two subjects, the employer (*principal*) and the worker (*agent*). In the first stage, the employer makes a job offer that consists of a binding wage offer w, and a desired, but non-binding effort level $\tilde{e}$. In the second stage, the worker can accept or reject the job offer. If he or she rejects the offer, the game is over. If the worker accepts, he or she chooses an effort level e, which is associated with costs $c(e)$. The higher the chosen effort e, the higher the payoff for the employer, but the higher the effort costs $c(e)$ for the worker. According to standard economic theory, a selfish worker, whose primary objective is to maximize his or her utility, will always choose the lowest effort level. The employer will anticipate this selfish behavior and offer the minimum wage possible (Fehr, Kirchsteiger, & Riedl, 1993; Fehr & Falk, 2002).

the *fair wage-effort hypothesis*, wages below the perceived fair wages may induce lower performance (Akerlof, 1982; Akerlof & Yellen, 1990).

1.2.2 Non-monetary incentives and their motivational properties

Non-monetary incentives can be classified as non-monetary tangible and non-monetary intangible incentives (Condly et al., 2003). Non-monetary tangible incentives are non-cash, but financial in nature, that is, they have a market value (Jeffrey, 2009). Examples include vacation trips, gift vouchers (Condly et al., 2003), luxury goods (Jeffrey, 2009), company cars, mobile phones and employer-provided meals (Hammermann & Mohnen, 2014a). Non-monetary intangible incentives relate to aspects such as positive performance appraisal, public praise (Condly et al., 2003), recognition (e.g., "employee of the week") and awards (Condly et al., 2003; Hansen & Weisbrod, 1972; Kosfeld & Neckermann, 2011), flexible working hours (Mas & Pallais, 2017; Wiswall & Zafar, 2018), titles or job autonomy (Aguinis, Joo, & Gottfredson, 2013).

Frey (2010) argues that non-monetary intangible incentives, such as awards, can signal recognition for outstanding performance. These awards must be visible and can, for example, be in the form of a public ceremony, a mention in the company magazine, or the intranet. Empirical evidence shows that a non-material award, such as a symbolic congratulatory card (Bradler, Dur, Neckermann, & Non, 2016; Kosfeld & Neckermann, 2011) or a medal (Chan, Frey, Gallus, & Torgler, 2014), has a positive impact on performance (Bradler et al., 2016; Chan et al., 2014; Kosfeld & Neckermann, 2011); this is especially the case when the award is publicly given, for instance, in a ceremony (Neckermann & Frey, 2013) or given exclusively to the best performers (Bradler et al., 2016). Furthermore, a public recognition program can lead to a significant increase in work attendance (Markham, Scott, & McKee, 2002). Neckermann, Cueni, and Frey (2014) point out that winning an award for voluntary activities can have a positive spillover effect on core performance, although this effect is short-lived. In addition, Grant and Gino (2010) show that even the simple verbal expression of gratitude can positively affect employees' effort. Status and social recognition are strong motivators for individuals (Besley & Ghatak, 2008; Ellingsen & Johannesson, 2007; Frey, 2007; Hubermann, Loch, & Önçüler, 2004; Kosfeld & Neckermann, 2011; Luthans, 2000; Markham et al., 2002). However, besides the positive effect of recognition, awards can also

have a negative effect due to demotivation caused by social comparison, in particular with regards to the weakest (Ashraf, Bandiera, & Lee, 2014). Moreover, Malmendier and Tate (2009) show that awards for Chief Executive Officers (CEOs), such as "CEO of the Year", may have a negative impact on their performance. After winning the award, CEOs indulge in activities that provide private benefits but do not increase firm value (e.g., writing books, consuming more perks, joining outside boards). Besides awards and praises, recent studies further analyze non-monetary incentives, such as flexible working hours; empirical evidence shows that people have a preference for flexible working schedules (Wiswall & Zafar, 2018) and value the option to work from home (Mas & Pallais, 2017). Furthermore, the meaning of work itself constitutes a crucial non-monetary aspect (Cassar & Meier, 2018); empirical evidence indicates that the feeling that one's work has meaning has a positive impact on performance (Ariely, Kamenica, & Prelec, 2008; Chandler & Kapelner, 2013; Kosfeld, Neckermann, & Yang, 2017).

Regarding non-monetary tangible incentives, the literature addresses (1) their use and associated advantages for companies, as well as (2) their impact on employee performance.

Several reasons compel companies to provide non-monetary incentives to their employees. One of the primary purposes of providing benefits seems to be the cost efficiency concern: companies have a comparative advantage over employees in buying benefits due to scale economies or tax treatment (Hammermann & Mohnen, 2014a; Oyer, 2008). Furthermore, by offering benefits, such as employer-provided meals or child care, companies aim to ease the effort costs of their employees (Hammermann & Mohnen, 2014a; Oyer, 2008). This is underlined by Rajan and Wulf (2006) who argue that non-monetary incentives, such as chauffeur service or financial counseling, are offered to enhance productivity, save time and reduce the effort costs of employees with leading responsibilities and a busy schedule. However, Oyer (2008) shows that the valuation of benefits differs across employees. It is associated with effort and costs for the employees to find a company that offers their preferred benefits. However, based on the reverse signaling approach introduced by Backes-Gellner and Tuor (2010), Hammermann and Mohnen (2014a) state that due to their high visibility, benefits can serve as signals for unobservable company characteristics, such as a good working environment, and thus help attract the target employees. Employees efficiently sort themselves to companies that provide the benefits that signal the characteristics they value most. Thus, this sorting effect reduces information asymmetry and counteracts

mismatching and job vacancies. Furthermore, their results suggest that benefits have a positive impact on work and wage satisfaction, as well as on employees' feeling of being acknowledged. Benefits seem to be a suitable means to reward employees for good work performance.

Regarding their effectiveness, non-monetary incentives are not beneficial from the point of view of standard economic theory. In particular, monetary incentives should always be better, or, at least no worse than non-monetary incentives due to their fungibility and option value. Cash can be used freely and everyone can buy what he or she wants according to his or her preferences (Jeffrey, 2009; List & Shogren, 1998; Waldfogel, 1993). However, empirical studies provide a more nuanced picture. Existing empirical research focuses on the effectiveness of non-monetary incentives either in tournaments, where incentives are dependent on performance (e.g., Hammermann & Mohnen, 2014b; Jeffrey, 2009), or in gift-exchange games, where incentives are given independently of performance (e.g., Kube et al., 2012; Mahmood & Zaman, 2010). Jeffrey (2009) investigates the impact of performance-related non-monetary incentives (massage vouchers) in a tournament setting and shows that performance is higher in the presence of non-monetary incentives than under monetary incentives of an equivalent value, although people state a preference for cash incentives. This behavioral inconsistency might be explained by justifiability concerns, as individuals may need to justify themselves if they purchase hedonic luxury goods on their own. In contrast, Hammermann and Mohnen (2014b) do not support these findings. They do not focus on justifiability concerns, but on the higher visibility of non-monetary incentives. Their experimental results suggest that monetary incentives induce a higher performance than non-monetary incentives. This is in line with the results of the meta-analysis conducted by Condly et al. (2003), which shows that monetary incentives have a higher impact than non-monetary incentives. However, Condly et al. (2003) note that their results are based on a limited number of studies on non-monetary incentives and, therefore, need to be interpreted with caution. Bareket-Bojmel et al. (2017) do not find significant differences in the effectiveness of performance-related monetary and non-monetary incentives. In addition, Presslee, Vance, and Webb (2013) shed light on the effects of non-monetary and monetary incentives on goal setting, goal commitment, and performance. Their results show that although non-monetary incentives lead individuals to choose less-challenging goals, subjects

seem more committed to achieving them. However, the average performance is higher under cash rewards due to their significant positive effect on the difficulty of the selected goals.

Considering the effectiveness of non-monetary performance-unrelated incentives in gift exchange games, Kube et al. (2012)'s controlled field experiment, in which workers are asked to catalog books at a university library, provides evidence that a non-monetary gift (thermos bottle) increases performance by 25%, whereas a cash reward of equivalent value has no significant positive impact on performance. Kube et al. (2012) suggest that the non-monetary gift might be seen as a kind act, which elicits positive reciprocal behavior. The time and effort the employer has invested in the gift matter to the employees (Baron & Kreps, 1999; Prendergast & Stole, 2001). In particular, the gift signals that the employer knows the employees' preferences (Prendergast & Stole, 2001). These results are confirmed by Mahmood and Zaman (2010) who analyze the impact of cash and non-cash rewards and the role of reciprocal behavior in a field experiment. According to Solnick and Hemenway (1996), tangible incentives have an emotional value for the recipient that exceeds their market value. Non-cash rewards can trigger emotional responses (Alonzo, 1996). Moreover, people think more frequently about non-monetary tangible incentives than about monetary incentives; this higher frequency of thoughts has a positive impact on performance (Jeffrey & Adomdza, 2010). Furthermore, in contrast with monetary incentives, which may crowd out intrinsic motivation, non-monetary incentives do not seem to discourage altruistic behavior. For instance, Lacetera and Macis (2010) show that only cash incentives lead to a decline in people's willingness to donate blood, whereas a voucher does not hinder prosocial behavior. In addition, Heyman and Ariely (2004) suggest that two kind of markets – a social market and a money market – determine the effectiveness of monetary and non-monetary incentives. In social markets, people act according to social norms and their effort is nearly independent of compensation levels and seems to be driven by altruistic motives. In contrast, in money markets, people base their effort decision on a simple cost-benefit analysis and a monotonic relationship exists between payment and performance.

Jeffrey and Shaffer (2007) identify four psychological approaches for explaining the motivational properties of non-monetary tangible incentives: *justifiability*, *separability*, *evaluability*, and *social reinforcement*. Tangible incentives are often seen as luxury items that appeal to people's hedonic desires, rather than fulfilling their utilitarian needs. Employees often feel guilty when they spend money on such items and feel the need to

justify themselves. However, if people earn these items as a reward for good performance, this guilt is relieved and employees do not feel the need to justify consumption, because the benefit has been not purchased (the *justifiability* argument). The *separability* argument is grounded in the mental accounting theory of Thaler (1999) and refers to the psychological fact that people tend to assign different types of incomes to different mental accounts instead of considering earnings collectively. Jeffrey and Shaffer (2007) state that due to its monetary nature, a cash bonus is merely seen as "more salary" and is evaluated relative to the base wage. Therefore, additional earnings have a diminishing marginal utility. In contrast, it is more likely that tangible incentives are not mentally added to the base wage and evaluated in isolation. Thus, they may have a more substantial impact than cash benefits. Furthermore, the concept of *evaluability* addresses the fact that non-monetary incentives allow subjects to cognitively alter the value of the benefit in both directions: they increase the value if the benefit seems to be attainable and decrease it if the benefit is perceived as being out of reach. Moreover, non-cash incentives have a trophy value and are highly visible to those in the recipient's social surroundings (the *social reinforcement* argument). By earning non-cash incentives, employees can enjoy the consumption value of the reward per se and also the recognition and acknowledgment of their performance by their social environment (Jeffrey & Shaffer, 2007). According to Ellingsen and Johannesson (2007), people have a desire for social esteem and enjoy earning respect and attention from their social environment. People strive for social recognition (Bandura, 1986; Stajkovic & Luthans, 2001; Stajkovic & Luthans, 2003). Moreover, as non-monetary incentives are highly visible, there is no need for employees to brag about their earnings. Nevertheless, it is much easier to talk about non-cash incentives then cash earnings (Jeffrey & Shaffer, 2007; Trachtmann, 1999).

1.3 Research questions

As outlined in Section 1.1, this dissertation aims to provide new insights regarding the effectiveness of monetary and non-monetary incentives from a behavioral economics perspective and to achieve a deeper understanding of their underlying mechanisms, in the interest of both researchers and human resources (HR) practitioners. This thesis comprises three separate studies, which address different topics in the incentive literature.

The initial focus of this thesis is on the impact of variable payment cuts on employee performance and satisfaction (***Chapter 2***). As outlined above, employees often receive variable bonus payments in addition to their monthly basic salary. These monetary incentives should help align the interests of employers and employees and improve the performance of the company (Prendergast, 1999). As the amount of variable payment – being linked to individual and company performance – is often not contractually stipulated, companies can more easily cut these variable payments than the fixed wages when they face financial distress (Cohen, 2018). For example, few years ago drastic bonus cuts were introduced at Deutsche Bank due to the financial crisis and the executive board as well as the middle management had to forego their bonus payments for 2016 (Frankfurter Allgemeine, 2017). So far, the effects of variable payment cuts have rarely been empirically analyzed in existing literature. According to standard theoretical predictions variable payment cuts have a negative impact on employee performance; more precisely, a monotonically decreasing relation between the level of the variable payment cut and the performance decrease is expected. However, the behavioral economic aspects, such as fairness concerns or negative reciprocity, may further affect employee behavior after a variable payment cut (cf. Kube et al., 2012; Lee & Rupp, 2007) and thus lead to behavior that differs from the predictions of standard theory.

To date, a considerable amount of research has dealt with the impact of fixed wage cuts on employee performance. Several studies show that fixed wage cuts have a negative and persistent impact on performance and point out that negative reciprocity plays an essential role in employment relations (Cohn, Fehr, Herrmann, & Schneider, 2014; Kube, Maréchal, & Puppe, 2013; Mas, 2006). However, there is also empirical evidence that does not support the existence of negative reciprocity in labor relations. If the reasons for the pay cut are regarded as comprehensive, as when a company is in financial distress, the pay cuts are perceived as “fair”; thus, the reduction of effort is short-lived (Lee & Rupp, 2007). Besides the literature on fixed wage cuts, initial research focuses on variable payment cuts showing that the complete removal of short-term incentives leads to a decrease in performance (Bareket-Bojmel et al., 2017). However, the major shortcoming in this area of research is that the effects of variable payment cuts of different levels on performance have not yet been analyzed empirically. Moreover, variable payment cuts not only affect employee performance but also their satisfaction. Previous studies indicate that cutting wages leads to

a deterioration in satisfaction (Kawaguchi & Ohtake, 2007; Pouliakas, 2010; Smith, 2015), poor work morale, and long-term damage to companies (Bewley, 2007). The extent to which different magnitudes of variable payment cuts affect satisfaction has not yet been analyzed in a laboratory environment. Thus, to address these gaps in research, *Chapter 2* considers the following research questions:

(1) *Do variable payment cuts of different levels have an impact on performance and satisfaction and, if so, to what extent*?

(2) *Do subjects facing a variable payment cut reduce their performance and satisfaction more than standard theory predicts*?

To this end, a real-effort laboratory experiment was conducted at the Technical University of Munich, where subjects were exposed to variable payment cuts of different levels. The experiment comprises five working periods, where subjects are expected to complete a task in each period, which consists of solving simple mathematical problems. This task allows for a precise measurement of work performance and quality (Hammermann & Mohnen, 2014b). The subjects' satisfaction with respect to the variable payments is revealed by the responses collected from each subject after the experiment. To identify significant treatment effects, the experimental data are analyzed using Wilcoxon rank-sum tests, as well as difference-in-differences (DiD) regression models.

Chapter 3 of this thesis analyzes the impact of monetary, non-monetary, and a combination of monetary and non-monetary incentives on performance and investigates which of these incentives leads to optimal performance. Non-monetary incentives are defined in this study as non-monetary tangible incentives with a measurable market value (Hammermann & Mohnen, 2014b; Jeffrey, 2009). As mentioned before, according to standard economic theory, monetary incentives are always better, or at least no worse than non-monetary incentives of equivalent value due to the option value of cash. The option value refers to the freedom offered by cash, which provides more opportunities to spend than a non-monetary incentive (Jeffrey, 2009; Waldfogel, 1993). Despite the existence of literature dealing with the effectiveness of monetary and non-monetary incentives, there is no clear evidence of the superiority of cash. On the one hand, some studies show that monetary incentives evoke higher performance than non-monetary incentives (Condly et al., 2003; Hammermann & Mohnen, 2014b). On the other hand, empirical evidence suggests that non-monetary

incentives induce higher performance (Jeffrey, 2009; Kube et al., 2012). Furthermore, some studies indicate that even when employees state a preference for monetary incentives, non-monetary incentives induce higher performance levels (Jeffrey, 2009; Kube et al., 2012). This study aims to analyze what kind of incentive – monetary or non-monetary – is superior and to provide an explanation for the mixed results in the literature regarding the effectiveness of these two types of incentives. Moreover, so far, no empirical evidence exists on the effectiveness of a combination of monetary and non-monetary incentives. These gaps in knowledge are summarized in the following research question:

(3) *Do monetary, non-monetary, or a combination of monetary and non-monetary (mixed) incentives have a positive impact on performance and which of these incentives induces the highest performance level?*

In addition, this study considers the impact of gender differences on the effectiveness of monetary, non-monetary, and mixed incentives. To date, a considerable amount of research has been devoted to gender differences in tournament settings and competition, showing that women dislike working in competitive environments, while men embrace competition (Datta Gupta, Poulsen, & Villeval, 2013; Dohmen & Falk, 2011; Masclet, Peterle, & Larribeau, 2015; Niederle & Vesterlund, 2007). Furthermore, the existing literature shows that women appreciate social aspects such as positive relationships, while men are more concerned with other job dimensions such as pay and promotions (Clark, 1997; Elizur, 1994). In addition, previous studies deal with gender differences regarding the impact of monetary incentives (Gneezy, Niederle, & Rustichini, 2003; Levitt, List, Neckermann, & Sadoff, 2016; Masclet et al., 2015; Niederle & Vesterlund, 2007) and discuss gender differences with regard to the motivational strengths of non-monetary tangible and intangible incentives in schools (Jalava, Schrøter Joensen, & Pellas, 2015; Levitt et al., 2016; Riener & Wagner, 2019). However, to the best of my knowledge, there is no empirical evidence analyzing which incentives – monetary, non-monetary, or mixed – are more effective in relation to gender differences. This lack of evidence motivates the fourth research question:

(4) *Are there gender differences in the effectiveness of monetary, non-monetary, and mixed incentives*?

This thesis addresses the third and fourth research questions through a laboratory experiment that considers one working period in which subjects are asked to solve mathematical

problems. Following Jeffrey (2009) and Hammermann and Mohnen (2014b), a tournament is implemented and participants can earn an additional prize on top of their fixed wage depending on their performance ranking. The experiment uses a between-subjects design with a control group (no prizes) and three different treatment groups, which differ in the prizes that subjects can earn (monetary prizes, non-monetary prizes, and a combination of monetary and non-monetary prizes). To analyze the effectiveness of these different incentives in a tournament setting, Wilcoxon rank-sum tests, as well as two-sided t-tests, ordinary least squares regressions (OLS), and appropriate post-hoc analysis of pairwise comparisons are performed.

Finally, ***Chapter 4*** considers the effectiveness of different non-monetary incentives of equal market value and a cafeteria-style incentive system, in which individuals are offered the choice between different non-monetary incentives. The use of non-cash incentives is widespread in today's business world (Incentive Research Foundation, 2017; Kauflin, 2017; Smith, Restle, & Stanger, 2015; Staufenbiel Institut, 2017; van Dyke, 2016; Zepelin, 2017). Existing research on the effectiveness of non-monetary incentives has generally found that non-monetary incentives enhance employee performance (Bareket-Bojmel et al., 2017; Jeffrey, 2009; Jeffrey & Shaffer, 2007; Kube et al., 2012; Mahmood & Zaman, 2010). However, research on the underlying psychological mechanisms of non-monetary incentives is still in its early stages (Ellingsen & Johannesson, 2007; Jeffrey & Shaffer, 2007; Prendergast & Stole, 2001; Solnick & Hemenway, 1996). Due to different tastes and preferences, non-monetary incentives of equal market value may not be equally effective for all individuals (Jeffrey & Shaffer, 2007; Jeffrey, 2009). However, the effectiveness of different non-monetary incentives of equivalent market value has not yet been examined in a research setting. Furthermore, although existing research recommends applying a cafeteria-style incentive system, in which employees can choose benefits that fit their individual needs best, to mitigate the problem of different preferences (Dzuranin, Randolph, & Stuart, 2013; Jeffrey & Shaffer, 2007), the effectiveness of a cafeteria-style system has not yet been empirically examined. This motivates the following research questions:

(5) *Are there any differences in the effectiveness of different non-monetary incentives of equivalent market value?*

(6) *Does a cafeteria-style incentive system have a positive impact on performance and does it induce a higher performance level compared to predetermined non-monetary incentives?*

These research questions are addressed using a laboratory experiment, which is an extension of the experiment described in *Chapter 3*. As in the previous experiment, participants are required to work on a mathematical task in a tournament setting, where they can earn an additional non-monetary prize on top of their fixed wage, depending on their performance rank. To compare the effectiveness of different non-monetary incentives, the experiment considers three treatments, each with predetermined non-monetary incentives of equal value. Furthermore, a treatment group with a cafeteria-style incentive system is also considered, in which subjects are offered the choice between three different non-monetary prizes. In addition to the laboratory experiment, a self-collected survey-based investigation is conducted to achieve a deeper understanding of the underlying psychological mechanisms and the effectiveness of a cafeteria-style incentive system. To analyze and compare the effectiveness of different non-monetary incentives of equal market value and a cafeteria-style incentive system, non-parametric Wilcoxon rank-sum tests, as well as Kruskal-Wallis tests, OLS regressions, and post-hoc analysis of pairwise comparisons are applied.

1.4 Methodology

"*Experimental economics is an 'exciting new development'.*"
(Samuelson & Nordhaus, 1992, p. 5)

This thesis addresses the research questions outlined in *Chapter 1.3* by using laboratory experiments in all three studies. It took a long time for the use of experimental methods to become established in economics (Friedman & Sunder, 1994). The first laboratory experiments in economics were conducted in the late 1940s (Falk & Heckmann, 2009). In the last few decades, the use of laboratory experiments in economics has witnessed steady growth (Falk & Fehr, 2003). The major breakthrough for experimental economics can be traced back to 2002, when Vernon Smith received the Nobel Prize together with Daniel Kahneman "[…] *for having established laboratory experiments as a tool in empirical economic analysis, especially in the study of alternative market mechanisms*" (Friedman & Cassar, 2004, p. 15). Since then, laboratory experiments have become an established method

in economics (Falk & Fehr, 2003). According to Smith (1994) there are several reasons why laboratory experiments are used in economics; they help test theories, discover reasons for a theory's failure, and identify empirical regularities, thereby serving as a foundation to build new theories. Furthermore, by using laboratory experiments, environments as well as institutions, can be compared and policy proposals can be evaluated. The laboratory can serve as a testing ground for institutional design. According to Falk and Fehr (2003) the most important advantage in conducting experiments in economics is control. The laboratory setting allows not only controlled variation but also controlling for the decision environment. No other method allows a comparable stringent control; the control possibilities in the laboratory environment exceed by far those available through other methods in the field (Falk & Fehr, 2003; Falk & Heckmann, 2009). Thus, laboratory experiments are characterized by a high degree of internal validity, which is understood as the degree to which the results can be attributed to the independent variables and causality can be inferred from the data (Falk & Fehr, 2003; Kühl, 2009; Loewenstein, 1999).[3] In this thesis, a laboratory experiment allows analyzing the relationship between different levels of variable payment cuts and performance (*Chapter 2*) as well as the effectiveness of different incentives (*Chapters 3* and *4*). For example, regarding the investigation of the impact of variable payment cuts (*Chapter 2*), it would be hard to investigate in the field whether a causal link exists between variable payment cuts and performance; this is because other variables (e.g., working conditions, relationship with employer and colleagues, working atmosphere) that cannot be controlled may have an influence on performance. However, under the controlled conditions of a laboratory experiment, it is possible to analyze the impact of variable payment cuts on performance when all other factors are held constant (Croson & Gächter, 2010; Falk & Fehr, 2003; Falk & Heckmann, 2009). Furthermore, a laboratory experiment allows one to observe and control variables that would not be observable otherwise. In the laboratory, performance of each individual can be measured with few measurement errors. Moreover, individual characteristics and preferences, such as ability, risk-aversion, and social preferences, can be observed (Dohmen & Falk, 2011). In this thesis, a laboratory experiment allows analyzing the impact of different incentives on individual performance

[3] Nevertheless, internal validity can only be ensured with proper experimental controls, a clean experimental design, and appropriate data analyses (Falk & Fehr, 2003; Friedman & Sunder, 1994).

accurately and observing the psychological mechanisms underlying the effectiveness of incentives.

However, the use of laboratory experiments does not only entail advantages. Objections exist regarding the use of experimental methods in economics. As the experimental participants are primarily students, it is often argued that a subject pool bias may exist, and the data collected in the laboratory may not be representative. Furthermore, sample sizes are often small and laboratory experiments do not capture all the essential conditions that may exist in reality. As a result, an indefinite amount of details will remain uncaptured in this artificial environment. Laboratory experiments do not depict the real situation, but only a simplified version of reality. Critics, therefore, claim that laboratory experiments lack realism and generalizability (Falk & Fehr, 2003; Falk & Heckmann, 2009; Friedman & Cassar, 2004; Friedman & Sunder, 1994). These objections regarding the external validity of experiments certainly need to be taken into account. However, there are studies showing that the differences that exist in the behavior of different subject pools (e.g., students vs. managers), are not fundamental and the qualitative patterns of behavior are quite alike (Cooper, Kagel, Lo, & Gu, 1999; Falk & Fehr, 2003; Fehr & List, 2004). Moreover, the simplified depiction of a situation within an experiment can also be an advantage, as it helps improve the understanding of the interplay between relevant variables (Falk & Fehr, 2003). Nevertheless, as suggested by Falk and Fehr (2003), we add more realism to the experimental setting by conducting real-effort experiments instead of chosen-effort experiments.[4] Despite the criticism regarding the external validity of laboratory experiments, this method seems suitable to analyze the research questions of this thesis, as it allows not only analyzing the

[4] In a *real-effort setting*, subjects work on a real task and their outcomes depend on their effort (Charness, Gneezy, & Henderson, 2018). Typical tasks include solving mazes (Gneezy et al., 2003), counting numbers (Pokorny, 2008), solving simple mathematical problems (Niederle & Vesterlund, 2007; Hammermann & Mohnen, 2014b), positioning sliders (Gill & Prowse, 2014), cracking walnuts (Fahr & Irlenbusch, 2000), or folding pieces of paper and stuffing envelopes (Falk & Ichino, 2006). The advantage of using a real-effort task is that it better reflects the working environment, e.g., there might be a variation of the cost of effort over time (Charness et al., 2018).
In a *chosen-effort* or *stated-effort setting*, experimental subjects choose their effort level from a menu (e.g., ranging from 1 to 10) with a corresponding list of costs; by choosing an effort level subjects incur certain monetary costs that increase with the level of effort chosen. To test a theoretical model using an experiment, it is crucial to control the main components of this theory. One essential aspect when testing theory is to control the cost of effort function. For example, when testing the LEN-model, the chosen-effort setting allows researchers to model a convex-cost function, meaning that the marginal cost of effort is increasing (Charness et al., 2018; Brüggen & Strobel, 2007). Several experimental studies use this *stated-effort* methodology; examples are Fehr et al. (1993), Fehr et al. (1997), Charness (2004), Fehr and Schmidt (2004), and Brandts and Cooper (2007) (Please note that this is neither an exhaustive list of studies using this method, nor a list of the most important ones).

causal links between variables but also precisely observing and studying human behavior (Dohmen & Falk, 2011; Falk & Fehr, 2003). Furthermore, company HR data measuring employee performance are mostly not available and have been used in the past only in exceptional cases. Notable examples are Knez and Simester (2001), who analyze the impact of the introduction of a new incentive scheme on employee performance at *Continental Airlines*, and Lazear (2000), who analyzes the effect of a switch from fixed pay to performance pay on the productivity of employees at *Safelite Glass Corporation*, a large auto glass company.

1.5 Contribution, main results, and structure of the thesis

This thesis makes several contributions, which are relevant for both researchers in the field of personnel economics, as well as managers and HR practitioners. Focusing on the effects of variable payment cuts and the effectiveness of monetary and non-monetary incentives, this thesis underlines the importance of understanding human behavior in an employment relationship that goes beyond the rational models.

Chapter 2 contributes to the existing literature by empirically analyzing how different levels of variable payment cuts affect effort provision and satisfaction. To the best of my knowledge, the effects of variable payment cuts of different levels on performance have not yet been analyzed empirically. Furthermore, the impact of variable payment cuts on satisfaction has not yet been examined in a laboratory environment.

The experimental results show that, contrary to the predictions of standard economic theory, variable payment cuts do not necessarily have a negative impact on performance; however, the level of the cut matters. Furthermore, subjects facing a high variable payment cut reduce their effort even more than standard theory predicts. The data suggest that behavioral economic considerations, such as fairness concerns as well as demotivation and goal setting aspects play an essential role in the employees' reaction to a variable payment cut. Moreover, a low cut leads to an increase in work quality. Whereas a cut of medium and high magnitude has no significant negative impact on work quality, a total cut leads to a significant quality reduction. Finally, a variable payment cut negatively affects pay satisfaction: once a certain

cut level is reached, satisfaction decreases to the same extent irrespective of the magnitude of the actual cut.

Chapter 3 analyzes the impact of monetary, non-monetary, and a combination of monetary and non-monetary incentives on performance. This experimental study contributes to the existing literature by providing an explanation for the mixed results in the existing experimental literature on the impact of these two types of incentives. This study also extends the existing literature by analyzing the impact of a combination of non-monetary and monetary incentives on performance. In addition, to the best of my knowledge, this is the first attempt to consider gender differences when investigating the effectiveness of monetary and non-monetary incentives.

The overall results suggest that no significant difference exists in performance in pursuit of monetary, non-monetary, and mixed incentives. However, gender-based differentiation reveals a more nuanced picture: while men's performance in response to monetary incentives is significantly higher than their performance in response to non-monetary incentives, women's performance is significantly higher in the presence of non-monetary incentives. Gender differences regarding the effectiveness of monetary and non-monetary incentives are not triggered by the perceived attractiveness of the non-monetary incentives but rather by differences between men and women in the feeling of appreciation and perceived performance pressure in the tournament setting. Furthermore, the results provide significant evidence that gender differences affect the effectiveness of monetary and non-monetary incentives, thus clarifying the mixed results regarding the impact of monetary and non-monetary incentives in the existing literature.

Chapter 4 contributes to the existing research by analyzing the effectiveness of different non-monetary incentives of equal market value in one setting. Furthermore, this study empirically investigates the effects of a cafeteria-style incentive system on performance, in which individuals are offered the choice between different non-monetary incentives. As such, this chapter emphasizes the importance of non-monetary incentives and the complexity of their underlying psychological mechanisms and motivational properties.

The experimental data indicate that non-monetary incentives of equal market value do not necessarily lead to the same performance level, but the attractiveness of prizes plays an important role in their effectiveness. However, a cafeteria-style incentive system has no

significant impact on performance, although it allows a selection of incentives in line with individual preferences. In addition, performance in response to a cafeteria-style system is not significantly higher than that in response to a preselected non-monetary incentive. In contrast, predetermined non-monetary incentives push individuals to a significantly higher performance level, especially when considering top performers. The poor performance of the cafeteria-style incentive system may be explained by the feeling of less appreciation, justification concerns, and effort costs arising from decision making.

Hence, the three studies presented in this thesis address relevant issues for human resource management (HRM) practices and incentive systems, and provide new research insights. Overall, the results indicate that not only self-interest but also socio-psychological aspects play an essential role in employment relationships and human decision making. Understanding the underlying mechanisms of incentives is crucial for implementing them effectively.

In the following sections, this thesis addresses the research questions presented in *Chapter 1.3*. The outline of this thesis is as follows: ***Chapter 2*** empirically investigates the impact of variable payment cuts on employee performance and satisfaction. ***Chapter 3*** analyzes the effectiveness of monetary, non-monetary, and a combination of monetary and non-monetary incentives. The objective of ***Chapter 4*** is to investigate the effectiveness of different non-monetary incentives of equal market value and a cafeteria-style incentive system. It should be noted that Chapters 2 through 4 represent a distinct scientific contribution on their own. These chapters are treated as independent studies, and each features an introduction, literature overview, methods, results, discussion, and conclusion section. Finally, ***Chapter 5*** summarizes the key findings of these studies and concludes by discussing their overall implications for both practice and future research.

2 Variable payment cuts and their impact on performance in the workplace: An experimental investigation[5]

2.1 Introduction

In today's business world, it is common for employees to receive variable bonus payments in addition to their basic wage. The purpose of the variable bonus payment is to incentivize employees and to ensure that they act in the best interest of the employer and contribute to improving the performance of the company (Holmstrom, 1979; Jensen & Meckling, 1976; Prendergast, 1999).

Variable payments have recently become increasingly significant in the compensation structure. The use of variable payments provides more flexibility to companies compared to permanent scheduled compensation increases, since they do not increase the fixed costs of the company (Cohen, 2018; McGregor, 2018). Moreover, in contrast to fixed basic wages, the amount of variable payments is often not contractually defined. It is therefore much easier for companies to cut a bonus payment than it is to do the same to a fixed monthly wage when the company experiences financial crises or changes in the economic environment. For example, the average bonus of bankers on Wall Street decreased by 9% in 2015 due to the financial crisis (Helmore, 2016; Moyer, 2016). Furthermore, at the beginning of 2017, Deutsche Bank announced drastic bonus cuts, with the majority of the executive employees not receiving a bonus at all for their performance in 2016 (Frankfurter Allgemeine, 2017; WirtschaftsWoche, 2017).

While bonus payment cuts seem to occur intermittently in the economy − especially during financial crises − scientists have rarely empirically analyzed their effects. Standard theoretical assumptions state that variable payment cuts lead to an equivalent performance

[5] This chapter is based on a working paper co-authored by Alwine Mohnen and Virginia Herbst. My contribution to the paper is summarized in Appendix D (signed by the authors in the examiners' copies of this dissertation). The working paper was presented at the 3rd International IAB/ZEW Workshop 2017 „Assessing the Impact of Human Resource Management Practices" in Nuremberg (IAB stands for *Institut für Arbeitsmarkt- und Berufsforschung*; and ZEW for *Zentrum für Europäische Wirtschaftsforschung*).

decrease, meaning that the higher the variable payment cut, the greater the performance decrease. Nevertheless, actual employee behavior may deviate from these predictions, as fairness concerns or negative reciprocity could play a role in employee reactions to bonus cuts (Kube et al., 2013; Lee & Rupp, 2007).

This study aims to investigate the impact of low, medium, high, and total variable payment cuts on performance and satisfaction. To this end, a real-effort experiment was conducted at a large German university. The experiment comprised five working periods. In each period, subjects were expected to complete a task, which consisted of solving simple mathematical problems. This means that work performance (the number of correctly solved mathematical problems) and work quality (the ratio of the number of correctly solved mathematical problems to total solved problems) were precisely measurable. Subjects received a fixed wage for each period and a variable payment for each correctly solved mathematical problem. During the experiment, subjects were exposed to variable payment cuts of different levels after the second working period to analyze whether subjects reacted to these different pay cut levels, and if so, to what extent. Furthermore, their satisfaction with the variable payments was revealed by the ex post responses of each subject after the experiment.

The results suggest that a variable payment cut does not necessarily have a negative impact on performance, but the level of the cut is crucial. In contrast to standard theoretical predictions, a low or medium variable payment cut does not significantly affect performance. However, a high and total variable payment cut both induce a significant decline in performance. Moreover, the performance of subjects facing a high variable payment cut reduces even more than standard theory predicts. Our results suggest that self-interest is not the only motivating factor for the subjects, but that fairness concerns as well as demotivation and goal setting aspects play an important role in their behavior after facing a variable payment cut. In addition, variable payment cuts have a negative impact on satisfaction. Once a certain cut level has been reached, satisfaction decreases to the same extent, irrespective of the cut level.

Existing research has mainly focused on the effects of fixed wage cuts on employee performance (Cohn et al., 2014; Greenberg, 1990; Kube et al., 2013) and on the impact of both fixed wage cuts and bonus cuts on satisfaction and work morale (Bewley, 2007; Kawaguchi & Ohtake, 2007; Pouliakas, 2010). Bareket-Bojmel et al. (2017), who analyze

the impact of short-term incentives on performance, shed further light on the impact of a total removal of these incentives. However, to the best of our knowledge, the effects of different levels of variable payment cuts on performance have not yet been analyzed empirically. Furthermore, the impact of the same on satisfaction has not yet been examined in terms of variable payment cuts of different levels. Therefore, this study contributes to the existing literature by analyzing how different levels of variable payment cuts affect effort provision and satisfaction in a controlled laboratory environment. Moreover, the study aims to provide insights for HR practitioners for designing payment cuts.

The remainder of this chapter is organized as follows: Sections 2.2 and 2.3 provide an overview of the existing literature and hypotheses, followed by a description of the experimental design and methods in Section 2.4. The results are outlined in Section 2.5, and Section 2.6 presents an in-depth discussion. Finally, Section 2.7 summarizes the key results, suggests possible paths for future research, and concludes by addressing the management implications of our findings.

2.2 Literature overview

While a considerable volume of empirical literature has addressed the impact of monetary incentives, particularly piece rates, on performance (Camerer et al., 1997; Gneezy & Rustichini, 2000; Lazear, 2000; Paarsch & Shearer, 2000; Pokorny, 2008), the effects of variable payment cuts have not yet been examined in depth. Until now, only a few researchers have investigated the impact of variable and fixed payment cuts on individual performance and satisfaction. By examining the impact of fixed wage cuts on workers' reciprocal behavior in a controlled field experiment, Kube et al. (2013) show that fixed wage cuts have a negative and persistent impact on work performance. Nevertheless, in contrast to their findings in terms of work quantity, they could not observe a decrease in work quality. In particular, their research results indicate that negative reciprocity plays an important role in employment relations. Further studies have also supported the importance of negative reciprocity in labor relations (Cohn et al., 2014; Mas, 2006). Moreover, Bareket-Bojmel et al. (2017) show that the total removal of monetary short term bonuses leads to a significant decrease in performance. However, in opposition to the aforementioned studies, Lee and Rupp (2007) only find limited empirical evidence that fixed wage reductions have a negative

effect on employees' effort provision. The findings of their natural field study show that significant and permanent pay cuts for U.S. commercial airline pilots only led to a temporary reduction of effort. According to Lee and Rupp (2007), employees' views on fairness provide a possible explanation for this finding, as most airlines were at or near bankruptcy when the cuts were made. Moreover, Chen and Horton (2016) show that workers react to wage cuts by reducing output if the justification for the cut is perceived as selfish and not reasonable. Bewley (2007) states that fairness is important in the context of labor relations, as this concept is crucial for good work morale and, therefore, an important factor in terms of influencing employee performance. Furthermore, DellaVigna, List, Malmendier, and Rao (2016) estimated employees' social preferences toward their employers in a field experiment by considering the employee's cost-of-effort function and the employer's payoff. While they exposed employees to pay raises as well as pay cuts, DellaVigna et al. (2016) discovered little evidence for reciprocal behavior. There is only a small negative decrease in output after an unexpected wage cut and employees' productivity remains almost constant.

The literature also suggests that fixed wage cuts, as well as bonus cuts, have a negative impact on satisfaction and lead to poor work morale. For example, Cohn et al.'s (2014) field experiment reveals that a general fixed wage cut not only results in lower effort provision by workers, but also in lower pay satisfaction when compared with employees whose wages had remained unchanged. Surveys conducted in Japan and Great Britain have come to similar conclusions on the effect of wage cuts on satisfaction (Kawaguchi & Ohtake, 2007; Pouliakas, 2010). Pouliakas's (2010), in particular, investigates the effect of incentive pay on job satisfaction, and presents evidence that workers are likely to experience a significant decrease in satisfaction if a bonus is withdrawn. Greenberg (1990) even provides evidence that temporary fixed wage cuts in manufacturing plants, which were not carefully explained to workers, can lead to feelings of frustration and severe reactions in terms of higher theft rates. Similar results are presented by Goette and Huffman (2005) who find that individuals show strong emotional reactions in response to fixed wage cuts.

However, to the best of our knowledge, the literature has not yet analyzed the effects different levels of variable payment cuts on performance and satisfaction. This study will address this gap by shedding light on how different levels of variable payment cuts affect work performance and satisfaction in a controlled laboratory setting.

2.3 Hypotheses

According to standard theoretical predictions of principal-agent models, higher incentives lead to higher performance, and lower incentives to lower performance (Prendergast, 1999; Spremann, 1987). In line with standard theory, we therefore state that individuals will exert lower effort after a variable payment cut. Furthermore, we predict a monotonically decreasing relationship between the amount of the variable payment cut and performance.

Hypothesis 1a: *A variable payment cut has a negative impact on performance.*

Hypothesis 1b: *The higher the variable payment cut, the higher the performance decrease.*

In addition, according to standard theoretical reasoning, we state the following:

Hypothesis 2a: *Following a variable payment cut, subjects exert the same effort compared with individuals who receive the same variable payment, but who were not exposed to a variable payment cut before.*

However, the consideration of behavioral economic aspects provides a different perspective. Kube et al. (2013) show in their field experiment that workers do not act purely out of self-interest. More specifically, they find that fixed wage cuts have a severe negative impact on individuals' performance, which indicates that negative reciprocity plays an essential role in the working environment. Cohn et al. (2014) support these results, and show by means of a field experiment that individuals reduce their effort by 15% after a fixed wage cut of 25%. In addition to issues of negative reciprocity, survey results suggest that employees' fairness perceptions concerning a cut have a strong impact on worker morale and effort (Bewley, 2007; Blinder & Choi, 1990; Smith, 2015). In their field experiment, Lee and Rupp (2007) provide further evidence for these findings, as concerns of fairness play an important role in employees' effort decisions when facing a wage cut. Based on a partial gift exchange game (Akerlof, 1982) and fair wage considerations (Kahneman, Knetsch, & Thaler, 1986), Lee and Rupp (2007) model employee effort as an increasing function of the wage offer itself, as well as of individual's fairness perception. We argue that exposure to a variable payment cut might be considered as unfair, and may, hence, lead to a performance decrease. Nevertheless, individuals still have an incentive to work following a variable payment cut (as long as they receive a variable payment); however, this incentive might be smaller than

standard theory predicts, as a variable payment cut may be perceived as unfair and thus, have a demotivating effect. Due to this, we suggest that a variable payment cut leads to lower performance when compared with the performance of individuals who received the same variable payment, but were not exposed to a variable payment cut before. Our hypothesis is formulated as follows:

Hypothesis 2b: *Following a variable payment cut, individuals exert lower effort compared with individuals who received the same variable payment, but who were not exposed to a variable payment cut before.*

Cohn et al. (2014) show that a reduction of fixed wages has a negative impact on work quantity, as well as on work quality. Greenberg's (1990) study supports these findings, and shows that wage cuts lead to feelings of frustration and detrimental work behavior. According to Spector and Fox (2002), negative emotions can lead to counterproductive work behavior, for example, incorrect execution of tasks. In contrast, the study of Kube et al. (2013) does not support these findings, as it reveals a higher work quality after a fixed wage cut. They suggest that wage reductions lead to a lower work effort, which is tantamount to a lower working speed; this, in turn, results in fewer mistakes and, consequently, in increased work quality. Although empirical evidence of the effects of fixed wage cuts on work quality is not clear, we state the following, based on the findings of Cohn et al. (2014) and Greenberg (1990):

Hypothesis 3: *A variable payment cut has a negative impact on work quality.*

In addition to the above, this paper not only addresses the effect of variable payment cuts on performance and work quality, but also on satisfaction. Pouliakas (2010) finds a significant decrease in satisfaction after the withdrawal of bonus payments. Furthermore, Kawaguchi and Ohtake (2007) report a robust finding that wage cuts lead to a deterioration of workers' pay satisfaction. Bewley (2007) states that wage reductions lead to poor work morale and long-term damage to companies. Moreover, Brown (2001) emphasizes the role of fairness concerns on satisfaction, and shows in her study that workers have a higher level of pay satisfaction if they believe that their compensation is fair. More specifically, Williams, McDaniel, and Nguyen (2006) state that the individuals' perceived fairness of their pay is directly related to pay satisfaction. Based on these empirical findings, we predict a decrease in satisfaction after a variable payment cut.

Hypothesis 4: *A variable payment cut leads to a decrease in satisfaction.*

2.4 Experimental design and data

To analyze the impact of variable payment cuts on workers' performance and satisfaction, we conducted a real-effort experiment using z-Tree (Fischbacher, 2007). We recruited participants using the ORSEE[6] online recruitment tool (Greiner, 2004) and assigned them randomly to one of nine treatment groups. The experiment consisted of five working periods, preceded by a test period to ensure that participants understood and became familiar with the task. The task was to solve simple mathematical problems consisting of two equations in which participants had to add or subtract three one-digit numbers. To calculate the final solution, the subjects had to subtract the lower result from the higher result of the single equations.[7] The advantage of this task is that it does not require any prior knowledge and is easy to understand. Moreover, the differences in performance and quality can be measured accurately (Hammermann & Mohnen, 2014b).

During the test period, participants had to solve five mathematical problems equal to the actual task. We used the time needed to solve these problems correctly (*testtime*) to measure ability for our analysis. Participants could continue the experiment only if they solved all five mathematical problems correctly. The performance in the test task did not affect the payment at the end of the experiment. Following the test period, the subjects could decide either to solve mathematical problems or to read articles of different genres (travel and discovery; technology and innovation; society, culture, and history; and politics and economics) during the five working periods. Each working period lasted five minutes. Subjects could switch between these two options at any time. Reading articles thereby only served as an outside option and did not affect the subject's payment. Nevertheless, Corgnet, Hernán-González, and Schniter (2015) suggest that providing a good alternative activity is important in a laboratory experiment to avoid having performance partly driven by the lack of "on-the-job" leisure activities and subjects working only out of boredom. Prior to each working round, participants received information on the payment scheme for the respective

[6] The abbreviation ORSEE stands for Online Recruitment System for Economic Experiments.
[7] The task is based on an experiment by Hammermann and Mohnen (2014b). For the detailed instructions of the mathematical task and an exemplary screen of the working period see Figure A1 and A2 in Appendix A.

working period, which we varied by treatment. Figure 2.1 provides an overview of the experiment timeline.

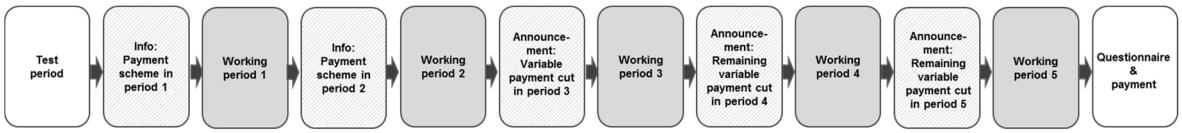
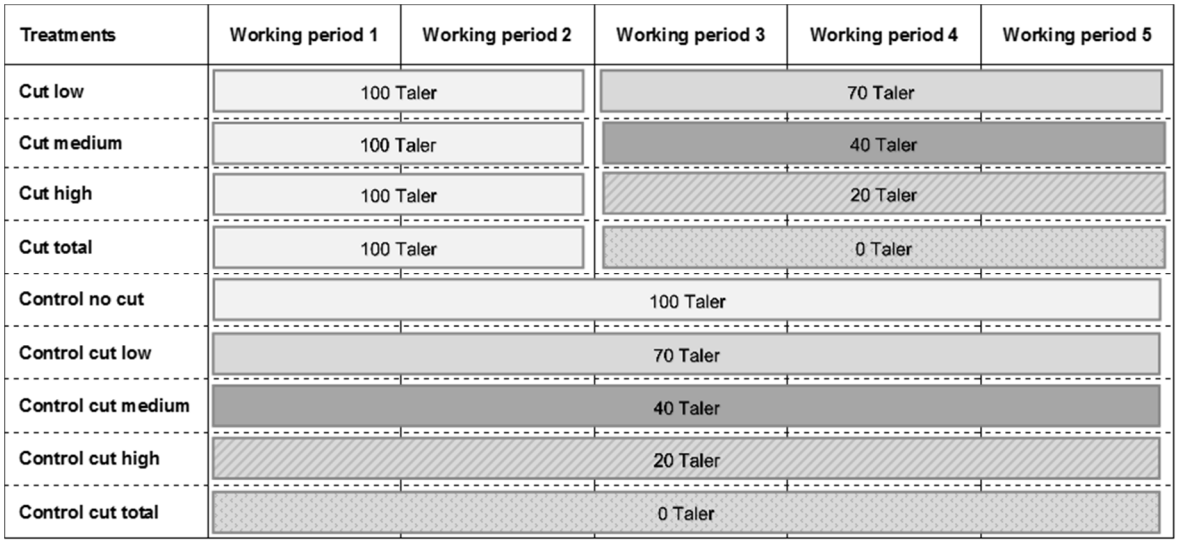


Figure 2.1: Experimental timeline (exemplary for cut treatments)

To answer our research questions, we had four treatments, representing the different levels of variable payment cut which took place after working period 2: *cut low*, *cut medium*, *cut high,* and *cut total*. In all four treatments, participants worked under a piece rate and a fixed payment of 3800 Taler (2050 Taler=1 euro) in each period. In the first two periods, the variable payment was 100 Taler for each correctly solved mathematical problem for all four treatments. After the second period, we announced different levels of variable payment cuts for period 3 in treatments *cut low*, *cut medium*, *cut high*, and *cut total*. Instead of 100 Taler, subjects received a lower variable payment for each correctly solved mathematical problem in period 3 (*cut low*: 70 Taler; *cut medium*: 40 Taler; *cut high*: 20 Taler; *cut total*: 0 Taler). Prior to working periods 4 and 5, we informed the subjects in these cut treatment groups again that their variable payment would remain cut (as in period 3) for the following period.

To examine the various incentive effects, we ran five control treatments. We did not expose subjects in the control treatments to a variable payment cut; regardless of the control treatment, the subjects received a fixed wage of 3800 Taler in each period. For subjects in the *control no cut* treatment, the variable payment of 100 Taler for every correctly solved mathematical problem remained unchanged over time. In the control treatments *control cut low*, *control cut medium*, and *control cut high*, the subjects received a variable payment of 70 Taler, 40 Taler, or 20 Taler for every correctly solved problem in each working period, respectively. Participants assigned to *control cut total* received no variable payment over all five working periods (for an overview of the variable payment scheme in the treatments see Figure 2.2). During the working periods, participants received no feedback about the number of correctly solved mathematical problems. At the end of the experiment, we asked the participants to fill out a short questionnaire and we announced the payout for the working periods.



Note: In addition to the variable payment, subjects in all treatments received a fixed wage of 3800 Taler for each of the five working periods, independent of their performance.

Figure 2.2: Variable payment scheme for the treatments over the five working periods

From June to October 2016, a total of 412 students at a large German university, participated in our experiment.[8] Their fields of study were mainly economics (31%), engineering (26%), and industrial engineering (11%). The average age of participants was 24 years (SD=3.62) and 39% were female. The participants were randomly assigned to one of the nine treatment groups (*cut low*: 46; *cut medium*: 46; *cut high*: 45; *cut total*: 45; *control no cut*: 45; *control cut low*: 44; *control cut medium*: 45; *control cut high*: 47; *control cut total*: 45). According to a Kruskal-Wallis test, there were no significant differences in gender and age between these groups (*gender*: p=0.297; *age*: p=0.590). The experiment lasted about 50 minutes with an average payment of 13.53 euros across all treatments.

2.5 Results

2.5.1 Work performance

As mentioned above, this study aims to analyze the impact of variable payment cuts on workers' performance measured by the number of correctly solved mathematical problems. A statistical summary of the main variables included in our analyses is reported in Table A1

[8] Four subjects had to be excluded as they filled in random numbers and did not seriously participate in the experiment (one in *cut high*, one in *control no cut*, and two in *control cut low*).

in Appendix A. The Kruskal-Wallis test indicated that no significant differences between treatments existed in ability, measured as the time needed to correctly solve the five mathematical problems in the test period (*testtime*, p=0.148). Furthermore, no differences in performance existed between treatments (*cut low*, *cut medium*, *cut high*, *cut total*, and *control no cut*) in periods 1 and 2, before the pay cut was announced (Kruskal-Wallis test, *score1*: p=0.375; *score2*: p=0.194). Given the simple nature of the mathematical problems, these findings suggest that our results were not driven by ability but by workers' effort. To measure the effect of variable payment cuts after the second working period in the short-term, we first conducted non-parametric tests and compared the absolute changes in performance between working periods 2 and 3 (*diffscore32=score3-score2*). This comparison of the average performance before and after the variable payment cut showed large differences within and between treatments (see Table 2.1).

Treatment	**Performance (mean) in working period 2** (before cut)	**Performance (mean) in working period 3** (after cut)	**Within treatment comparison** Delta (period 3-period 2)[a]	**Comparison with *control no cut*** Delta (period 3-period 2)[b]
Cut low	35.09	37.04	+1.95*** (0.002)	+0.93 (0.261)
Cut medium	33.76	32.93	-0.83 (0.874)	-1.85 (0.266)
Cut high	31.98	24.69	-7.29** (0.021)	-8.31*** (0.005)
Cut total	33.91	19.93	-13.98*** (0.000)	-15.00*** (0.000)
Control no cut	35.09	36.11	+1.02** (0.049)	-

Note: [a] Wilcoxon signed-rank test; [b] Wilcoxon rank-sum test; p-values in parentheses; * p<0.10, ** p<0.05, *** p< 0.01.

Table 2.1: Performance before and after the variable payment cut

Those who faced a low variable payment cut, did not decrease, but rather increased their performance. Subjects' output in the *cut low* treatment group was 5.56% higher in period 3 compared to period 2. Compared with the control group *control no cut*, the *diffscore32* was slightly higher, although this difference was not statistically significant. A medium cut led to a slight decrease in performance (-2.46%), not significantly lower compared to the *control no cut* group. By contrast, when workers faced a high or total variable payment cut, performance decreased tremendously on average by 22.80% and 41.23%, respectively, between periods 2 and 3, and the results were highly significant. Furthermore, the differences in scores between periods 2 and 3 significantly varied between the treatments and the benchmark treatment *control no cut*. Thus, our results partially contradict hypothesis 1a.

Using the balanced panel data structure, we performed difference-in-differences (DiD) regression analyses to assess the robustness of the non-parametric test results and to analyze in more detail the impact of variable payment cuts over working periods 1 to 5. Figure 2.3 illustrates the average work performance across the five working periods under the five different treatment conditions and shows that the effects remained stable across the periods.

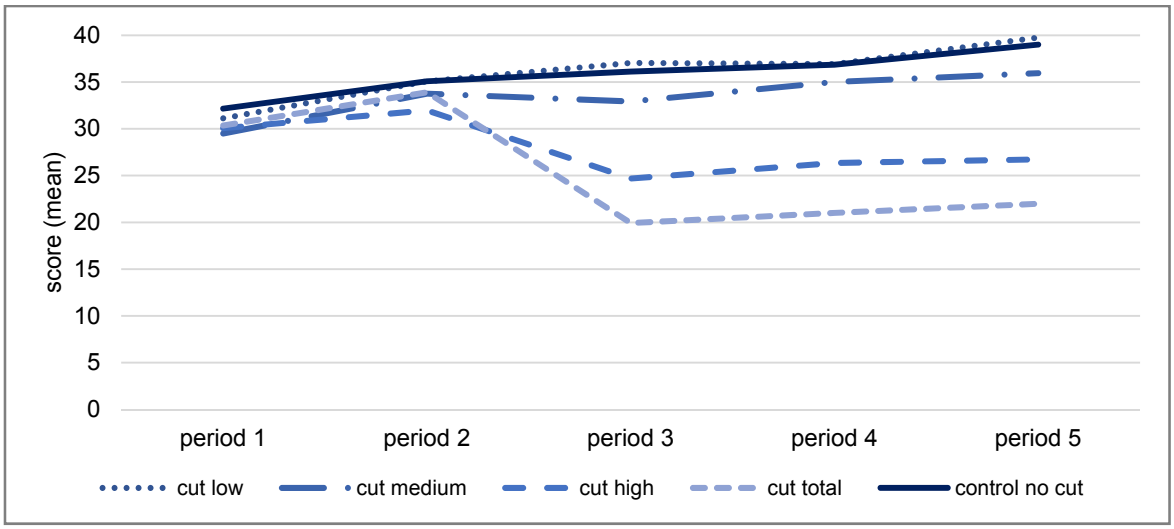


Figure 2.3: Performance (*score*) over the five working periods and the different treatments

For our difference-in-differences regression analyses, we thus averaged the data over the observed periods as follows: we grouped periods 1 and 2 − before the variable payment cut − as the pre-cut phase, and we grouped periods 3 to 5 − after the variable payment cut − as the post-cut phase. We included the variable *post-cut* in the model as a dummy variable that equaled 1, if the observation was from the post-cut phase, and 0, if it was from the pre-cut phase. The coefficient of the variable *post-cut* thus measured the difference between the average performance in the periods after the variable payment cut and that in the periods before the said cut. Furthermore, the binary variables for the various treatments *cut low*, *cut medium*, *cut high*, and *cut total* indicated whether a subject was part of one of these four treatment groups. The dummy variable for the benchmark treatment *control no cut* served as the reference category. We included the interaction terms *cut low x post-cut*, *cut medium x post-cut*, *cut high x post-cut*, and *cut total x post-cut* in our model. These interaction terms acted as the *difference-in-differences estimators* of interest, also called the *average treatment effects* (Wooldridge, 2016); they were utilized to measure the differences in performance before and after the variable payment cut for the respective treatments relative to the benchmark treatment *control no cut*. In addition, we included the independent variable

competition (motivation due to competitive thinking) in our model, which reflected the statement "I wanted to be better than other participants", where agreement or disagreement was measured on a seven-point Likert scale. Finally, we inserted control variables for personal characteristics (*age* and *gender*) and ability (*testtime*).

The results of the DiD regression models, estimated with robust standard errors adjusted for clustering at the individual level, are reported in Table 2.2. Overall, they support the findings of the non-parametric tests.

	Model (1) *score*	**Model (2)** *score*	**Model (3)** *outtime*
Cut low x post-cut	1.108	1.108	-3.861
	(1.43)	(1.43)	(-1.11)
Cut medium x post-cut	-0.689	-0.689	9.929
	(-0.62)	(-0.62)	(1.35)
Cut high x post-cut	-8.793***	-8.793***	61.66***
	(-4.67)	(-4.67)	(3.84)
Cut total x post-cut	-14.84***	-14.84***	100.7***
	(-6.10)	(-6.09)	(5.65)
Post-cut	3.696***	3.696***	2.889
	(6.60)	(6.59)	(1.20)
Cut low	-0.514	0.236	-4.351
	(-0.33)	(0.16)	(-1.15)
Cut medium	-1.992	-0.791	-2.086
	(-1.14)	(-0.48)	(-0.56)
Cut high	-2.600	-0.467	-5.444
	(-1.61)	(-0.31)	(-1.21)
Cut total	-1.489	0.0302	-3.042
	(-0.87)	(0.02)	(-0.80)
Age (in years)		-0.220	1.197
		(-1.38)	(1.47)
Gender (=1 if female)		-0.377	-16.76***
		(-0.33)	(-2.77)
Testtime (in sec.)		-0.0278**	-0.0729**
		(-2.33)	(-2.27)
Competition (1-7)		1.835***	-8.266***
		(6.12)	(-4.75)
_cons	33.62***	32.37***	28.17
	(26.74)	(7.52)	(1.29)
R^2	0.179	0.288	0.297
adj. R^2	0.173	0.279	0.289
N	1135	1135	1135

Note: Robust t-statistics corrected for clusters on individual level in parentheses; ***, **, * indicate the statistical significance at the 1%, 5%, and 10% level, respectively; the dependent variable is the number of correctly solved mathematical problems (*score*) in Model (1) and (2), and the time subjects spent reading articles (*outtime*) in Model (3).

Table 2.2: Estimation results of the DiD regression models on work performance

Model (2) shows that a low cut led to a performance increase. Nevertheless, this effect was quite small and not statistically significant (p=0.154). The subjects' performance in the *cut medium* group did not significantly decrease after the variable payment cut compared to the

control no cut group (p=0.534). In contrast, a high cut and a total cut both led to a detrimental decrease in performance, significant at the 1% level (*cut high*: p=0.000; *cut total*: p=0.000). Furthermore, post-estimation tests revealed that the higher the variable payment cut, the greater the decrease in scores (Wald test, *cut low* vs. *cut medium*: F(1,226)=2.71, p=0.101; *cut medium* vs. *cut high*: F(1,226)=15.85, p=0.000; *cut high* vs. *cut total*: F(1,226)=4.13, p=0.043). Excepting subjects exposed to a low variable payment cut, whose performance did not decrease, results are in line with hypothesis 1b.

Some explanatory variables had a significant impact on performance as well (see Model (2)). For instance, motivation evoked by perceived competition was positively correlated with performance (p=0.000). Moreover, the lower workers' ability, the worse their performance (p=0.021).

As an additional robustness check, we addressed the time subjects spent reading articles during the working period (*outtime*) as an alternative performance measure. The results are reported in Model (3) in Table 2.2 and provide further support for our conclusions.

Our main findings are summarized as follows:

Result 1a. *While a low cut and a medium cut do not significantly affect performance, a high or total cut leads to a significant decrease in performance.*

Result 1b. *Excepting subjects exposed to a low variable payment cut, the data suggest that the higher the variable payment cut, the higher the decrease in performance.*

These results indicate that beyond a certain magnitude of cut a significant negative relationship between cut level and performance is observed. To analyze in detail the incentive level after a variable payment cut, we investigated whether subjects in the cut treatments exerted the same effort after a cut, when compared to those who received the same variable payment, but had not been exposed to a cut before (see Figure A3 in Appendix A). According to standard economic assumptions these effort levels, due to the same incentive levels, should be the same. Table 2.3 shows that in the *cut low* group the number of correctly solved mathematical problems after the variable payment cut was significantly higher than in treatment *control cut low* (contradicts hypotheses 2a and 2b).

Score (mean)	Cut low	Control cut low	Delta (%)[a]	Cut medium	Control cut medium	Delta (%)[a]	Cut high	Control cut high	Delta (%)[a]	Cut total	Control cut total	Delta (%)[a]
Pre-cut phase	33.11	29.78	11.18%** (0.026)	31.63	33.12	-4.50% (0.277)	31.02	28.35	9.42% (0.243)	32.13	21.46	49.72%*** (0.000)
Post-cut phase	37.91	33.89	11.86%* (0.075)	34.64	36.11	-4.07% (0.448)	25.93	32.08	-19.17%** (0.023)	20.99	18.53	13.28% (0.466)

Note: [a] p-values for the Wilcoxon rank-sum test in parentheses (results remain stable using two-sided t-tests); * significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level; *score*: number of correctly solved mathematical problems.

Table 2.3: Performance comparison between the treatment groups and respective control groups in the pre-cut phase and post-cut phase

In contrast, there were no differences in performance between the *cut medium* and *control cut medium* group in the post-cut phase (in line with hypothesis 2a). Performance in the *cut high* group was significantly lower than in the respective control group in the periods after the cut, as stated in hypothesis 2b. Moreover, subjects' performance in the post-cut phase in the *cut total* group and the respective control group, where no variable payment was paid at all, was not statistically different (which supports hypothesis 2a).

Result 2. *Subjects in the cut low treatment group exert a higher effort compared to subjects in the control cut low group that receive the same variable payment, but have not been exposed to a variable payment cut before. In contrast, workers in the cut high group exert a lower effort compared to subjects in the respective control group that receive the same variable payment, but have not been exposed to a variable payment cut before.*

2.5.2 Work quality

We complemented the insights on work performance with evidence on work quality. Following a similar approach as Kube et al. (2013) and Hennig-Schmidt, Sadrieh, and Rockenbach (2010), we measured work quality as the ratio of correctly solved mathematical problems to the total number of problems worked. Table A1 in Appendix A provides an overview of the average work quality (in percentage) across the five working periods for the different treatments. The data show that under the treatment *cut low*, only a slight decrease in work quality is observable between periods 2 and 3 (-0.30%) after the variable payment cut was announced, while work quality in the *cut medium* and *cut high* treatment groups decreased by 2.16% and 4.61%, respectively. Subjects in the *cut total* treatment showed a

decrease in their work quality of 15.68%. For comparison, work quality in our benchmark treatment *control no cut* decreased by 1.21%. In order to analyze whether the decrease in work quality was affected by the variable payment cut, rather than just evoked, for example, due to the fatigue induced by the repetitive task and the monotony of the work (Gonzalez, Best, Healy, Kole, & Bourne, 2011; Healy, Kole, Buck-Gengler, & Bourne, 2004), we estimated a difference-in-differences regression model with *quality* as our dependent variable (Table 2.4).

	Model (1) *quality*	**Model (2)** *quality*
Cut low x post-cut	0.0215*	0.0215*
	(1.76)	(1.75)
Cut medium x post-cut	-0.00241	-0.00300
	(-0.10)	(-0.13)
Cut high x post-cut	-0.0318	-0.0356
	(-1.27)	(-1.37)
Cut total x post-cut	-0.0539**	-0.0558**
	(-1.98)	(-2.02)
Post-cut	-0.0274**	-0.0274**
	(-2.47)	(-2.47)
Cut low	-0.00563	-0.000790
	(-0.72)	(-0.09)
Cut medium	-0.0213**	-0.0204**
	(-2.42)	(-2.18)
Cut high	-0.00676	0.000372
	(-0.73)	(0.04)
Cut total	-0.0164	-0.0151
	(-1.60)	(-1.35)
Age (in years)		0.000172
		(0.10)
Gender (=1 if female)		0.0263**
		(2.20)
Testtime (in sec.)		-0.0000277
		(-0.49)
Competition (1-7)		0.00836**
		(2.06)
_cons	0.951***	0.899***
	(149.86)	(16.32)
R^2	0.058	0.082
adj. R^2	0.050	0.071
N	1081	1081

Note: Robust t-statistics corrected for clusters on individual level in parentheses; ***, **,* indicate the statistical significance at the 1%, 5%, and 10% level, respectively; the dependent variable is the work quality (*quality)*, measured as the ratio of correctly solved mathematical problems to the total number of problems worked.

Table 2.4: Estimation results of the DiD regression models on work quality

The estimation results of the difference-in-differences regression (see Table 2.4, Model (2)) show that those who experienced a low cut exhibited a significant increase in their work quality compared to the *control no cut* group after facing a variable payment cut (p=0.081). The decrease in work quality for the *cut medium* treatment group did not differ significantly

compared to that in the *control no cut* group (p=0.898). Moreover, a high cut had a negative impact on work quality when compared to the *control no cut* condition, although this difference did not reach the 10% significance level (p=0.172). Subjects who experienced a total cut showed a significant decrease in their work quality (p=0.045). Therefore, in contrast to hypothesis 3, which states that a variable payment cut harms work quality, these results suggest that a variable payment cut does not necessarily lead to a decrease in work quality, but the level of the cut matters.

Moreover, we found empirical evidence that the work quality of those who experienced a variable payment cut, did not significantly differ in the periods after the cut compared to the work quality of subjects who received the same variable payment but were not exposed to a variable payment cut before (see Table 2.5).

Work quality in percentage	Cut low	Control cut low	p-value	Cut medium	Control cut medium	p-value	Cut high	Control cut high	p-value	Cut total	Control cut total	p-value
Pre-cut phase	94.54	93.13	0.289	92.97	94.76	0.009	94.43	94.25	0.888	93.47	93.11	0.522
Post-cut phase	93.95	93.75	0.749	91.15	93.31	0.921	89.69	92.86	0.498	88.20	87.80	0.785

Note: p-values reported according to a Wilcoxon rank-sum test.

Table 2.5: Comparison of work quality between the cut treatments and respective control treatments in the pre-cut phase and post-cut phase

We summarize our findings as follows:

Result 3. *A low variable payment cut has a positive impact on work quality. While a medium or high variable payment cut has no significant impact on work quality, a total cut leads to a significant decrease in work quality.*

2.5.3 Satisfaction with variable payment

Finally, we complemented the insights from the experiment with survey data on satisfaction with variable payments. Subjects were asked ex post the working periods about their satisfaction with the variable payment for each of the five working periods. The level of satisfaction had to be rated on a Likert scale from 0, indicating complete dissatisfaction, to 10, corresponding to complete satisfaction. Our data show that subjects' stated satisfaction with variable payment significantly decreased following a variable payment cut between

periods 2 and 3 in the short-term: whereas a low cut led to a decrease in the amount of 29.27%, a medium cut evoked a decrease in the amount of 58.14%. Furthermore, following a high or total cut, subjects' stated satisfaction decreased on average by 64.43% and 67.12%, respectively (see Table 2.6 and Figure A4 in Appendix A).

Treatment	**Satisfaction with variable payment in period 2 (mean)**	**Satisfaction with variable payment in period 3 (mean)**	**Within comparison** Delta (period 3-period 2)[a]	**Comparison with *control no cut*** Delta (period 3-period 2)[b]
Cut low	6.15	4.35	-1.80*** (0.000)	-1.63*** (0.000)
Cut medium	7.00	2.93	-4.07*** (0.000)	-3.90*** (0.000)
Cut high	7.00	2.49	-4.51*** (0.000)	-4.34*** (0.000)
Cut total	6.69	2.20	-4.49*** (0.000)	-4.32*** (0.000)
Control no cut	5.13	4.96	-0.17** (0.019)	-

Note: [a] Wilcoxon signed-rank tests; [b] Wilcoxon rank-sum tests; p-values in parentheses: ***, **, * indicate the statistical significance at the 1%, 5%, and 10% level, respectively.

Table 2.6: Comparison of the subjects' average satisfaction with variable payment before and after the announcement of the variable payment cut

Compared to the benchmark treatment *control no cut*, this decrease in satisfaction with variable payment from period 2 to 3 was highly significant for all cut treatments, which is in line with our hypothesis 4 (see Table 2.6). Considering all periods 1 to 5, a difference-in-differences regression confirmed these results. The coefficients of the DiD estimators (*cut low x post-cut*, *cut medium x post-cut*, *cut high x post-cut*, *cut total x post-cut*) all had a negative sign and were significant at the 1% level (see Table 2.7, Model (2)).

Comparing the stated satisfaction in the different cut treatments, our data indicate that, compared to treatment *cut low*, subjects in the *cut medium*, *cut high*, and *cut total* group had a significantly higher decrease in satisfaction after the variable payment cut was announced (Wilcoxon rank-sum test, *cut low* vs. *cut medium*: p=0.000; *cut low* vs. *cut high*: p=0.000; *cut low* vs. *cut total*: p=0.000). According to a Kruskal-Wallis test, the decrease in satisfaction from round 2 to 3 did not differ between the treatments *cut medium*, *cut high*, and *cut total* (p=0.561). This result shows that if the cut is equal or higher than 60%, the satisfaction decreases to the same extent, irrespective of the cut level. Pairwise post-estimation tests after the difference-in-differences regression confirmed this result (see Table 2.7).

	Model (1) *satisfaction*	**Model (2)** *satisfaction*
Cut low x post-cut	-1.619***	-1.619***
	(-4.85)	(-4.84)
Cut medium x post-cut	-3.920***	-3.920***
	(-11.56)	(-11.53)
Cut high x post-cut	-4.337***	-4.337***
	(-11.92)	(-11.89)
Cut total x post-cut	-4.374***	-4.374***
	(-9.13)	(-9.11)
Post-cut	-0.344**	-0.344**
	(-2.20)	(-2.20)
Cut low	0.986*	1.049*
	(1.73)	(1.85)
Cut medium	1.844***	1.852***
	(3.66)	(3.69)
Cut high	1.833***	1.893***
	(3.77)	(3.82)
Cut total	1.611***	1.639***
	(2.90)	(2.90)
Age (in years)		0.00920
		(0.19)
Gender (=1 if female)		0.227
		(0.72)
Testtime (in sec.)		0.000905
		(0.43)
Competition (1-7)		0.0782
		(0.92)
_cons	5.167***	4.387***
	(13.35)	(3.32)
R^2	0.342	0.346
adj. R^2	0.337	0.338
N	1135	1135
Post-estimation tests (Wald tests):		
Cut medium x post-cut vs. *cut high x post-cut*:		p=0.351
Cut medium x post-cut vs. *cut total x post-cut:*		p=0.406
Cut high x post-cut vs. *cut total x post-cut*:		p=0.947

Note: Robust t-statistics corrected for clusters on individual level in parentheses; ***, **,* indicate the statistical significance at the 1%, 5%, and 10% level, respectively; the dependent variable is the satisfaction with variable payment (*satisfaction*).

Table 2.7: Estimation results of the DiD regression models on satisfaction with variable payment

In addition, we found empirical evidence that, although subjects received the same variable payment in the post-cut phase, those who experienced a medium or high cut before, were significantly less satisfied than those who had not previously experienced a cut. There were no significant differences in satisfaction between treatment group *cut low* and *control cut low* in the periods 3 to 5 (see Table 2.8).

Thus, the following conclusion can be drawn:

Result 4. *After a variable payment cut, satisfaction with the variable payment decreases significantly. From a certain cut level upwards, satisfaction decreases to the*

same extent, irrespective of the cut level. Compared with subjects receiving the same variable payment, but who have not been exposed to a variable payment cut before, subjects in the cut medium and cut high treatment have a significantly lower pay satisfaction.

Satisfaction with variable payment (mean)	Cut low	Control cut low	Delta[a]	Cut medium	Control cut medium	Delta[a]	Cut high	Control cut high	Delta[a]
Pre-cut phase	6.15	4.85	1.30** (0.028)	7.01	4.26	2.75*** (0.000)	7.00	3.71	3.29*** (0.000)
Post-cut phase	4.19	4.85	-0.66 (0.337)	2.75	4.01	-1.26** (0.047)	2.32	3.82	-1.50*** (0.007)

Note: [a] Wilcoxon rank-sum test for testing mean differences, p-values in parentheses, results remain stable using two-sided t–tests with equal variances; ***, **, * indicate the statistical significance at the 1%, 5%, and 10% level, respectively.

Table 2.8: Comparison of satisfaction with variable payment (mean) between the treatment groups and respective control groups in the pre-cut phase and post-cut phase

2.6 Discussion

Our experimental results show that a variable payment cut does not necessarily lead to a performance decrease, but that the cut level is decisive. To obtain further insights into subjects' effort decisions following a variable payment cut, we investigated the questions on the perception of fairness regarding the variable payment and the motivational aspects ex post the experiment.

Empirical evidence indicates that fairness concerns might play an important role in subjects' choice of effort following wage reductions (Blinder & Choi, 1990; Lee & Rupp, 2007; Smith, 2015). To assess subjects' fairness perceptions, they were asked ex post the experiment how fair they perceived their variable payment to be in each working period. The level of perceived fairness was rated on a Likert scale from 0, indicating complete unfairness to 10, corresponding to complete fairness. The self-reported results indicated that a variable payment cut had a significantly negative impact on the valuation of fairness with respect to the variable payment in all cut treatments, compared to the benchmark treatment *control no cut*. The higher the decrease in the fairness perception of the variable payment between the pre-cut and post-cut phase, the higher the decrease in performance between the said phases in the treatments *cut low*, *cut medium*, *cut high*, and *cut total* (Spearman rank correlation coefficient, $\rho(182)=0.203$, $p=0.006$). While subjects' fairness perception regarding the

variable payment only slightly decreased from the pre-cut phase to the post-cut phase for the *cut low* group, subjects in the *cut medium*, *cut high*, and *cut total* group had a significantly higher decrease in fairness perception compared to the *cut low* treatment group. Furthermore, the decrease in fairness perception was higher in the *cut high* and *cut total* group compared to the *cut medium* group, although this difference was only significant for *cut high* (see Table 2.9 and Table A1 in Appendix A).

	Fairness perception of variable payments		**Within treatment comparison:**
	Pre-cut phase	**Post-cut phase**	**Delta (=post-cut phase – pre-cut phase)[a]**
Cut low	5.74 (2.72)	4.19 (2.47)	-1.55*** (p=0.000)
Cut medium	6.30 (2.20)	2.96 (2.17)	-3.34*** (p=0.000)
Cut high	6.76 (1.93)	2.39 (2.12)	-4.37*** (p=0.000)
Cut total	6.47 (2.32)	2.41 (2.66)	-4.06*** (p=0.000)
Control no cut	5.11 (2.41)	4.84 (2.40)	-0.27** (p=0.032)

Pairwise comparisons of perceived fairness decrease (delta) using Wilcoxon rank-sum tests (p-values reported):			
Control no cut vs. cut low:	p=0.000***	Cut low vs. cut high:	p=0.000***
Control no cut vs. cut medium:	p=0.000***	Cut low vs. cut total:	p=0.000***
Control no cut vs. cut high:	p=0.000***	Cut medium vs. cut high:	p=0.030**
Control no cut vs. cut total:	p=0.000***	Cut medium vs. cut total:	p=0.232
Cut low vs. cut medium:	p=0.000***	Cut high vs. cut total:	p=0.493

Note: Mean values reported, standard deviations in parentheses; [a] Wilcoxon signed-rank test ***, **, * indicate the statistical significance at the 1%, 5%, and 10% level, respectively.

Table 2.9: Pairwise comparison of the difference of perceived fairness of the variable payment between the pre-cut phase and post-cut phase

The decrease in fairness perception might further evoke the decrease in subjects' satisfaction with variable payment (Brown, 2001; Williams et al., 2006). According to Gneezy and Rustichini (2000) a piece rate that is too small can be considered as insulting in itself.

Besides the consideration of fairness perception, subjects were asked if they were discouraged by the variable payment cut (measured on a seven-point Likert scale). The Spearman rank correlation coefficient indicates that the more demotivated subjects were, the higher the performance decrease after facing a variable payment cut ($\rho(182)=-0.298$, $p=0.000$). These results agree with the findings of Bewley (2007) who states that wage cuts lead to a lowering of work morale. A comparison of subjects' statements between the different cut treatments shows that subjects in the *cut low* and *cut medium* treatment stated to feel not as demotivated as subjects in the *cut high* or *cut total* treatment (see Table 2.10).

	Cut low	Cut medium	Cut high
Cut medium	+28.38%** (0.025)	-	-
Cut high	+41.22%** (0.013)	+10.00% (0.404)	-
Cut total	+62.16%*** (0.000)	+26.32%** (0.016)	+14.83% (0.211)

Note: ***, **,* indicate the significance at the 1%, 5%, and 10% level using the Wilcoxon rank-sum test; demotivation was rated on a Likert scale with 1, indicating "not demotivated at all" and 7, corresponding to "completely demotivated". Stated mean demotivation in *cut low* was 2.96 (SD=1.79); in *cut medium* 3.80 (SD=1.85); in *cut high* 4.18 (SD=2.35); and in *cut total* 4.80 (SD=2.30).

Table 2.10: Pairwise comparison of stated demotivation (mean) following a variable payment cut

Furthermore, our data reveal that fairness perception and demotivation might be potential drivers for the lower incentive effect after the variable payment cut in the *cut high* treatment when compared with the respective control treatment *control cut high*, in which subjects received the same variable payment, but were not exposed to a variable payment cut before. On average, subjects in the *cut high* group stated to perceive the variable payment in periods 3 to 5 significantly less fair than those in the *control cut high* group, although they received the same variable payment (Wilcoxon rank-sum test, p=0.003).

Although subjects in the *cut low* and *cut medium* group stated to be less demotivated and their perceived fairness perception decreased less than those in the *cut high* and *cut total* group after facing a variable payment cut, this does not explain why the subjects did not decrease their performance after a variable payment cut.

Goal setting theory could possibly explain this behavioral pattern (Locke, Bryan, & Kendall, 1968; Locke & Latham, 2002; Locke, Shaw, Saari, & Latham, 1981). Following Locke et al. (1981) a goal is something that an individual is trying to achieve. Goal setting can improve performance, especially when goal attainment is linked to monetary rewards. Nevertheless, if the target is too distant and seemingly unachievable for the individual, people get demotivated and decrease their performance (Locke & Latham, 2013). In our experimental setting, we asked subjects ex post if they had a certain wage in mind that they wanted to earn during their participation in this experiment and, if so, after which period they felt that they had achieved this target. As a result, 93% of the subjects in the cut treatment groups, who stated ex post that they had expected to earn a certain wage (target), believed they had not reached their target after working period 2, when their variable payment was cut. Individuals in the *cut low* and *cut medium* treatment groups, might have decided to exert more effort

after the announcement of the variable payment cut in order to still reach their target. Despite the cut, their goal still seemed to be achievable; however, in order to reach it, subjects had to work more, as they received a lower variable payment for each correctly solved mathematical problem compared with working period 2. If the cut was too large (*cut high*, *cut total*), subjects might conclude that their target could not be achieved; they therefore became discouraged and decreased their performance (Locke & Latham, 2013). These considerations are underlined by statements made by the subjects ex post the experiment. Subjects in the *cut low* and the *cut medium* treatment groups stated that they significantly exerted more effort after the variable payment cut in order to maximize their payoff, in comparison with the statements of subjects in the other cut treatments (Wilcoxon rank-sum test, *cut low* vs. *cut high*: p=0.001; *cut low* vs. *cut total*: p=0.001; *cut medium* vs. *cut high*: p=0.055; *cut medium* vs. *cut total*: p=0.036). Moreover, as mentioned above, subjects in the *cut high* and *cut total* treatment groups stated that the variable payment cuts resulted in a higher level of demotivation, compared with subjects in the in the *cut low* or *cut medium* treatment group.

2.7 Conclusion

This study experimentally analyzed the impact of variable payment cuts on performance and satisfaction in a real-effort setting. Based on our analysis, the following key findings are made. First, our data reveal that a variable payment cut does not necessarily lead to a performance decrease, but, instead, it depends on the level of the cut. While a low cut and a medium cut both do not significantly affect performance, a high or total cut leads to a significant decrease in performance. Moreover, subjects facing a high variable payment cut reduce their performance significantly more than standard theory predicts. Our results suggest that self-interest is not the only motivating factor for subjects, but that fairness concerns, demotivation and goal setting aspects play an important role in their behavior. Second, a low variable payment cut leads to a significant increase in work quality. While a medium or high cut has no significant negative impact on work quality, a total cut leads to a significant decrease in work quality. Finally, a variable payment cut has a negative impact on pay satisfaction. From a certain cut level upwards, the decrease in satisfaction is equal, irrespective of the actual cut level.

However, there are some limitations to the generalizability of our results. First, we did not provide any explanation for the variable payment cut to participants in our experimental setting. Thus, further work should be undertaken to explore whether, and to what extent, an indication of reasons for the variable payment cut might have an impact on subjects' behavior, performance, and satisfaction. Workers might perceive variable payment cuts differently if they better comprehend why the cut is necessary, for example to guarantee jobs during a recession (Chen & Horton, 2016; Greenberg, 1990; Kube et al., 2013; Lee & Rupp, 2007). Second, it would be worth deepening the understanding of the underlying psychological mechanisms and behavioral patterns that play a role in the effort decision process, following a variable payment cut.

Despite the constraints of our experimental setting, this study contributes to existing studies on monetary incentives by analyzing how subjects react to variable payment cuts. To the best of our knowledge it is the first empirical study on the impact of different levels of variable payment cuts on performance in a laboratory environment. The findings strongly indicate that subjects' actual behavior can deviate from standard theoretical predictions. Behavioral economic considerations, such as fairness concerns and demotivation, as well as goal setting, seem to play an essential role in subjects' effort decisions after following a variable payment cut.

In summary, our study indicates that a variable payment cut does not necessarily lead to a performance decrease, but that the cut level is decisive. Thus, the size of the cut should be determined with care. Moreover, HR practitioners should consider that variable payment cuts seem to evoke a feeling of demotivation and lead to dissatisfaction, and these reactions are sensitive to the cut level. From a certain cut level upwards, satisfaction seems to decrease to the same extent without regard to the extent of cut. Therefore, variable payment cuts should be implemented with considerable caution in companies.

3 Cash, non-cash, or mix? Gender matters! The impact of monetary, non-monetary, and mixed incentives on performance[9]

3.1 Introduction

Cash is king? According to standard economic theory, a monetary incentive is always better − or at least no worse − than a non-monetary incentive of equal market value, due to the option value of cash (Jeffrey, 2009; Waldfogel, 1993). It is often difficult for a company to determine the preferences of each individual employee and to choose the most suitable non-monetary incentives. It is reasonable to assume that companies may occasionally choose inappropriate material incentives that do not match employees' preferences. Therefore, employees would be better off receiving cash incentives, enabling them to buy a benefit which maximizes their individual utility (Jeffrey, 2009). However, although the use of non-monetary benefits is not reasonable from a neoclassical point of view, it is a widespread phenomenon within companies (Kauflin, 2017; Zepelin, 2017).

This study investigates the effects of monetary, non-monetary, and a combination of monetary and non-monetary (mixed) incentives on performance. To this end, we conducted a laboratory experiment with four different treatments (*monetary*, *nonmonetary*, *mix*, and *control*) and implemented a tournament, in which participants could earn an extra prize in addition to their fixed wage, according to their ranking position based on their performance. The additional prize depends upon the treatment group: while subjects in the *monetary* treatment group received cash prizes, subjects in the *nonmonetary* treatment group received non-monetary prizes (*Lindt* chocolates), and those in the *mix* treatment group a combination of non-monetary and monetary prizes (cash and *Lindt* chocolates). The task consisted of

[9] This chapter is based on a working paper co-authored by Alwine Mohnen. My contribution to the paper is summarized in Appendix D (signed by the authors in the examiners' copies of this dissertation). The working paper was presented at the *21st Colloquium on Personnel Economics (COPE)* 2018 in Munich, at the *Bavarian Micro Day* 2018 in Munich, and at the *XIX. Symposium zur Ökonomischen Analyse der Unternehmung* 2018 in Frankfurt organized by the German Economic Association of Business Administration (GEABA). As part of the *GEABA* conference 2018, an earlier version of this working paper was published in the *Discussion Paper Series in Economics and Management*.

solving simple mathematical problems and the number of correctly solved mathematical problems served as a performance measure.

Our experimental data indicate that overall, there is no significant difference in performance between the groups in pursuit of monetary, non-monetary, or a combination of non-monetary and monetary incentives, although performance in pursuit of mixed incentives is slightly higher than that of the pure non-monetary or pure monetary incentive groups. However, upon considering the genders separately, a different picture is revealed: while men's performance in response to monetary incentives is significantly higher than their performance in response to non-monetary incentives, women's performance is significantly higher in pursuit of non-monetary incentives. Furthermore, our results show that these gender differences regarding the impact of monetary and non-monetary incentives on performance are not evoked by the prize attractiveness.

To date, economic literature has focused mainly on monetary incentives. Monetary incentives are considered to be powerful incentives that are suitable for enhancing employees' performance (Condly et al., 2003; Jenkins et al., 1998; Prendergast, 1999). However, several existing empirical studies show that monetary incentives do not always enhance performance, and can possibly have a detrimental effect. This negative effect on performance is often ascribed to a motivation crowding-out effect or to the existence of reference-dependent preferences (Camerer et al., 1997; Deci & Ryan, 2002; Frey, 1997; Frey & Jegen, 2001; Gneezy & Rustichini, 2000; Mellström & Johannesson, 2008; Pokorny, 2008; Titmuss, 1970). With regard to non-monetary incentives, existing empirical research focuses on their impact on performance either within the context of gift-exchange games, where incentives are given independently of performance (Kube et al., 2012; Mahmood & Zaman, 2010), or within tournaments, where incentives are directly related to performance (Hammermann & Mohnen, 2014b; Jeffrey, 2009). However, empirical research regarding the effectiveness and underlying psychological mechanisms of non-monetary incentives is still in its early stages and there is currently no clear evidence on whether monetary or non-monetary incentives are more effective.

The contribution of this study to the existing literature is threefold. First, the effects of monetary and non-monetary incentives in a tournament setting are analyzed to gain a clearer perspective and to provide an explanation for the mixed results in the existing experimental

literature pertaining to the effectiveness and superiority of these two types of incentives. Second, the study endeavors to extend existing literature by analyzing the effects of a combination of non-monetary and monetary incentives on performance in tournaments. Third, the study contributes to the literature by analyzing gender differences concerning the impact of non-monetary, monetary, and mixed incentives. To the best of our knowledge, neither mixed incentives nor gender differences regarding the impact of different kinds of incentives on performance have been analyzed before.

The remainder of this chapter is organized as follows: Section 3.2 presents an overview of the existing literature and our hypotheses, followed by a description of the experimental design in Section 3.3, while Section 3.4 presents the results. Finally, Section 3.5 provides an in-depth discussion and Section 3.6 a concluding summary, containing the management implications of our findings as well as limitations and directions for future research.

3.2 Literature overview and hypotheses

According to standard economic theory, monetary incentives should always be better, or at least no worse, than non-monetary incentives are (Jeffrey, 2009; Waldfogel, 1993). Nevertheless, several empirical studies show, to the contrary, that non-monetary incentives can have a stronger positive effect on performance than do monetary incentives of equivalent value. In a controlled field experiment, in which workers had to catalog books at a university library, Kube et al. (2012) analyze the impact of performance-unrelated incentives on performance. Their study reveals that people show a 25% higher performance when they receive a non-monetary gift (thermos bottle), whereas the cash gift of the equivalent value has no significant impact on their productivity, although people state a preference for receiving the cash gift over the non-cash gift. The authors suggest that people might see the non-monetary gift as an act of generosity by the employer, who evidently invested time and effort into the gift, which therefore elicits positive reciprocal behavior. Mahmood and Zaman (2010), who analyze the impact of cash and non-cash gifts and the role of reciprocal behavior in a field experiment, confirm these results. Their study shows that non-monetary gifts elicit more reciprocal behavior and result in significantly better performance than cash gifts do. Furthermore, Lacetera and Macis (2010) show in an experimental study that while cash has a detrimental effect on the willingness to donate blood, non-monetary incentives like a

voucher do not backfire on pro-social activities. Jeffrey and Shaffer (2007) identify four key psychological concepts that explain the motivational power and effectiveness of non-monetary tangible incentives: *justifiability*, *social reinforcement*, *separability* (based on Thaler (1999)), and *evaluability* (see Jeffrey and Shaffer (2007) for an overview). Moreover, there is empirical evidence that people think more often about non-monetary tangible incentives than about monetary incentives; this higher frequency of thoughts positively affects performance (Jeffrey & Adomdza, 2010). According to Heyman and Ariely (2004), the type of market, whether monetary or social, determines the relationship between payment and effort. In the former case effort seems to stem from reciprocal motives and subjects determine their effort based on a simple cost-benefit analysis; in the latter case, where non-monetary incentives rather than cash are at stake, effort seems to stem from altruistic motives.

Besides the impact of performance-unrelated monetary and non-monetary incentives, existing literature analyzes the effects of performance-related incentives. Kelly, Presslee, and Webb (2017) examine a repeated tournament setting and find that while cash and non-cash incentives did not evoke different performance levels in the first tournament, the first tournament losers in the second tournament performed better in pursuit of non-cash incentives than in pursuit of cash incentives. Thus, non-cash incentives had a higher performance effect than cash incentives in the second tournament. Moreover, by means of a laboratory experiment, Jeffrey (2009) investigates the motivational power of tangible non-cash incentives. His results show that adults performed better in pursuit of non-monetary incentives than in pursuit of monetary incentives of the same value, although people stated a preference for the monetary incentive. He explains this result in terms of justification concerns, as people might have to justify the purchase of hedonic luxury goods. However, Hammermann and Mohnen's (2014b) experimental study does not support these findings. The authors analyze work performance in competitions in pursuit of non-monetary and monetary prizes. In contrast to Jeffrey (2009), they do not focus on justification concerns, but on the higher visibility of non-monetary incentives. Their results show that subjects performed better in pursuit of monetary incentives than in pursuit of non-monetary incentives. This is in line with results of Condly et al. (2003) who show by means of a meta-analytic review that money has a higher impact on performance than non-monetary tangible incentives do. Nevertheless, Condly et al. (2003) also remark that the generalizability of their

findings is limited as they are based on a small number of studies considering non-monetary incentives and the actual market values of the non-monetary incentives used in their meta-analysis could not be determined. In contrast to the aforementioned studies, Bareket-Bojmel et al. (2017) do not find a significant difference in the impact of non-monetary and monetary incentives on productivity in their field study.

Although the empirical evidence on whether non-monetary or monetary incentives are superior is not clear, we can assume, based on Jeffrey (2009), that performance is higher in pursuit of non-monetary incentives. Kube et al. (2012) suggest that a non-monetary gift is considered a kind act and is therefore positively reciprocated with higher performance (Kube et al., 2012). The massage vouchers used in Jeffrey (2009) as non-monetary incentives might have had a high reward and gift character. However, the pens used in Hammermann and Mohnen (2014b) might have rather been seen as a useful item, instead of an honorable reward, thus reducing the positive effect of rewarding good performance. Furthermore, following Jeffrey and Shaffer (2007), we argue that – in contrast to monetary incentives – non-monetary incentives have motivational properties in themselves (in addition to the market value of the incentive), as they are highly visible (the *social reinforcement* argument) and can be evaluated independently of other income (the *separability* argument). Employees not only have the utility of the incentive per se, but they also enjoy the recognition and acknowledgement of their performance within their social environment. While cash bonuses are typically invisible to others and people usually avoid bragging about monetary rewards, non-monetary rewards have a trophy value and are highly visible, which fosters social communication of the employee's strong performance (Jeffrey & Shaffer, 2007). Individuals strive for social esteem and recognition (Bandura, 1986; Ellingsen & Johannesson, 2007; Stajkovic & Luthans, 2003) and empirical evidence shows that non-monetary intangible incentives, like symbolic awards have a positive impact on performance (Kosfeld & Neckermann, 2011; Neckermann et al., 2014; Neckermann & Frey, 2013). Regarding the *separability* argument, Thaler's (1999) mental accounting theory suggests that people have different mental accounts and do not consider their income collectively; thus, they cognitively divide different components of their income and value them separately in different mental accounts (Jeffrey & Shaffer, 2007). Jeffrey and Shaffer (2007) emphasize that additional earnings might have a diminishing marginal utility for the employee, as he or she will mentally combine these earnings with the base salary and thus evaluate them relative

to this base salary. In contrast, employees will evaluate non-monetary incentives separately from the base salary and hold them in a separate mental account. Choi and Presslee (2016)'s experimental results support this argument. Their results suggest that subjects are more likely to categorize non-monetary incentives separately from salary than monetary incentives, and that subjects' performance is higher when they categorize performance-related pay separately from salary. Following the results of previous studies, we define our hypotheses as follows:

Hypothesis 1. *Monetary, non-monetary, and mixed incentives have a positive impact on performance.*

Hypothesis 2. *Performance is higher in pursuit of non-monetary incentives than in pursuit of monetary incentives.*

Furthermore, Kube et al. (2012) show a statistically higher effect of an artistically folded and wrapped monetary gift compared to a pure monetary gift in their field study. The treatment effect of the folded monetary gift was even slightly higher than that of the non-monetary gift, although this difference was not significant. They thus suggest that the effort and time an employer invests into the gift matters. We assume that by mixing monetary and non-monetary incentives, the employer can combine the benefits of monetary incentives (option value) with the benefits of non-monetary incentives (e.g., visibility within the social environment, social reinforcement, appreciation of the employer as he invested time and effort). Moreover, mixed incentives may suit both subjects with preferences for monetary incentives as well as those with preferences for non-monetary incentives. Therefore, mixed incentives should evoke a higher overall performance than either pure non-monetary or pure monetary incentives will. Nevertheless, to the best of our knowledge, there is no existing literature on the effectiveness of mixed incentives so far. Following our reasoning leads to our third hypothesis:

Hypothesis 3. *Performance is higher in pursuit of mixed incentives than in pursuit of either pure monetary incentives or pure non-monetary incentives.*

Moreover, this study addresses the possible role that gender differences play in the impact of monetary, non-monetary, and mixed incentives on performance, which, to the best of our knowledge, has not yet been analyzed in existing research. To date, existing literature deals

with gender differences in various incentive schemes. Masclet et al. (2015) show that women's performance is significantly higher than men's performance in fixed payment schemes and suggest that these results are mainly attributable to the fact that women are more intrinsically motivated than men. Moreover, existing studies show that there are no significant differences in performance in simple piece rate schemes (Gneezy et al., 2003; Niederle & Vesterlund, 2007). In contrast, Levitt et al. (2016) show that under low financial incentives, boys have a significantly higher performance compared to girls, whereas in the non-financial treatment, where they can earn a trophy, there are no differences in performance. Special focus has been devoted to gender differences in tournament schemes and competition. Several studies show that women are reluctant to work in competitive environments and shy away from competition, while men embrace competitive environments (Buser, Niederle, & Oosterbeek, 2014; Datta Gupta et al., 2013; Dohmen & Falk, 2011; Masclet et al., 2015; Niederle & Vesterlund, 2007). While women perform worse than men in competitive environments, their performance is similar in non-competitive environments (Gneezy et al., 2003; Gneezy & Rustichini, 2004; Shurchkov, 2012). Overconfidence, preferences for competition (Niederle & Vesterlund, 2007), and risk aversion (Dohmen & Falk, 2011) are often cited as reasons for these gender differences; men tend to be more overconfident (Niederle & Vesterlund, 2007) and less risk averse (Croson & Gneezy, 2009; Dohmen & Falk, 2011; Eckel & Grossmann, 2002) than women. Furthermore, while women falter under performance pressure, men do very well (Azmat, Calsamiglia, & Iriberri, 2016; Shurchkov, 2012). Following existing literature, we note that individuals are more focused on output, when monetary incentives are at stake; this might lead to higher competitiveness, as individuals strive for monetary prizes (Hammermann & Mohnen, 2014b; Vohs, Mead, & Goode, 2008). According to Heyman and Ariely (2004), money affects subjects' perceptions and results in a shift from a social to a money market. In contrast, non-monetary prizes might reframe the competitive market as a more social market, thereby weakening the competitiveness of a tournament. Thus, women might feel more comfortable in a competition where non-monetary incentives are at stake and therefore perform better than in a competition with monetary prizes. In contrast, men seek competition and therefore perform better when monetary prizes are at stake, as they become more persistent in pursuing them.

Possible gender differences in the effectiveness of monetary and non-monetary incentives might be due not only to different reactions to perceived competitiveness and performance pressure, but also to the feeling of appreciation. The existing literature reveals that non-monetary benefits can address employees' needs for acknowledgement and are a suitable means to reward employees for good performance and to show respect. This may therefore have a positive effect on employees' effort choice (Ellingsen & Johannesson, 2007; Hammermann & Mohnen, 2014a; Kube et al., 2012). Nevertheless, the most appropriate kind of incentive to reward and acknowledge employees might differ between men and women. Several studies show that extrinsic job dimensions such as pay and prospects for promotion are of high importance for men, while women value the more social aspects such as a positive relationship with managers (Clark, 1997; Elizur, 1994). Jalava et al. (2015) show by means of a field study in schools that females respond strongly to non-monetary incentives like a certificate reward as they attach great importance to reflected appreciation. When the employer invests time in seeking and buying a prize, female employees may perceive the prize as much more personal than a pure monetary prize; thus, the non-monetary prize may signal more appreciation and evoke a higher degree of positive reciprocity and performance (Kube et al., 2012; Prendergast & Stole, 2001). We hence suggest that women may feel more appreciated by non-monetary incentives than by monetary incentives, and may therefore perform better in pursuit of such incentives than in pursuit of monetary or mixed incentives. Conversely, as men are more concerned with pay, we assume that the reverse will be true for men.

We posit the following hypotheses regarding gender differences:

Hypothesis 4. *Men's performance is higher in pursuit of monetary incentives than in pursuit of non-monetary or mixed incentives.*

Hypothesis 5. *Women's performance is higher in pursuit of non-monetary incentives than in pursuit of monetary or mixed incentives.*

3.3 Experimental design and data

In order to analyze the effects of monetary, non-monetary, and mixed incentives, we conducted a real-effort experiment using z-Tree (Fischbacher, 2007). Participants were

recruited using the online recruitment system ORSEE (Greiner, 2004) and were randomly assigned to one of four treatment groups. The experiment consisted of one working period and the task was to solve simple mathematical problems. Each mathematical problem contained two equations, each consisting of three one digit numbers, which had to be added or subtracted. In order to calculate the final solution subjects had to subtract the lower result from the higher result of the single equations.[10] In order to ensure that participants understood the task, the working period was preceded by a test period, in which subjects had to solve five mathematical problems. Participants were only permitted to continue, if they answered all five mathematical problems correctly. The time, subjects needed to solve the test equations correctly (*testtime*) served as an ability checker. Before participants started the working period, they were informed that they were to work for 15 minutes and that they could decide to either solve mathematical problems or to read articles of different genres (society, culture, travel, economics, science, and technology). Subjects were allowed to switch between these two options at any time. Reading articles only served as an outside option and was not relevant for performance rankings. However, according to Corgnet et al. (2015) it is important to offer an outside option to avoid performance triggered by boredom and people only working because of a lack of desirable alternatives in the laboratory. Furthermore, participants were informed that they would receive a fixed wage of 10 euros, and that they could earn an additional prize, which depended on their performance ranking position.[11] Following Hammermann and Mohnen (2014b) and Jeffrey (2009), we implemented a tournament with four performance ranking groups to avoid a middle group. Each session included 22 subjects. The first ranking group consisted of the subject who performed best, the second group included the ranking positions 2-8, the third group 9-18, and the worst group comprised the ranking positions 19-22. Performance was measured by the number of correctly solved mathematical problems (*score*). If two participants had the same score, then the ratio of correctly solved problems to overall worked problems decided the ranking. If this indicator was still equal, the ranking position was decided by chance.

Altogether, we conducted four treatments, which differed in the prizes subjects could earn according to their relative performance ranking. In our benchmark treatment *control* subjects

[10] The task is based on Hammermann & Mohnen (2014b). According to existing research, there are no gender differences in the ability to solve simple mathematical problems (Hyde, Fennema, & Lamon, 1990; Niederle & Vesterlund, 2007).

[11] We decided to implement a tournament instead of a piece rate scheme, as companies commonly reward employees on the basis of their performance relative to the performance of other employees (e.g., sales staff).

only received a fixed wage of 10 euros without any additional prize, but they were informed of their relative position afterwards. In treatments *monetary*, *nonmonetary*, and *mix* participants were able to earn an additional prize according to their relative performance position on top of their fixed wage of 10 euros. The value of prizes increased with performance and ranking position. In treatment *monetary* the best performing subject received 10 euros, subjects of the second performance ranking group 5 euros, subjects of the third group 2.50 euros, and those of the worst group received no prize. The prize available for the best in treatment *nonmonetary* was a large box of *Lindt* chocolates (*Lindt Pralinés Hochfein*) worth 10 euros, subjects in the second group received a medium-sized box of *Lindt* chocolates (*Lindt Mini Pralinés* 100g) worth 5 euros, and subjects in the third group received a small box of *Lindt* chocolates (*Lindt Mini Pralinés* 44g) worth 2.50 euros. As in treatment *monetary*, subjects in the fourth group received no prize. Finally, in order to analyze the effects of a combination of non-monetary and monetary incentives, we implemented the treatment *mix*, where subjects received a combination of non-monetary and monetary incentives based on their performance ranking: 5 euros + a medium-sized box of *Lindt* chocolates (*Lindt Mini Pralinés* 100g), worth 5 euros; 2.50 euros and a small box of Lindt chocolates (*Lindt Mini Pralinés* 44g), worth 2.50 euros; and 1.25 euros + two *Lindt* chocolates (*Fioretto*), worth 1.25 euros; or 0 euros, respectively (see Figure 3.1 for an overview of treatments and prizes).

Ranking / Treatment	Ranking position 1 (prize value: 10 €)	Ranking positions 2-8 (prize value: 5 €)	Ranking positions 9-18 (prize value: 2.50 €)	Ranking positions 19-22 (prize value: 0 €)
Monetary	10 €	5 €	2.50 €	No prize
Nonmonetary	 *Lindt Pralinés Hochfein*	 *Lindt Mini Pralinés (100g)*	 *Lindt Mini Pralinés (44g)*	No prize
Mix	5 € +  *Lindt Mini Pralinés (100g)*	2.50 € +  *Lindt Mini Pralinés (44g)*	1.25 € +  *2 Lindt Fioretto*	No prize
Control	No prize	No prize	No prize	No prize

Figure 3.1: Overview of prizes

Prior to the working period, subjects in the treatments *nonmonetary* and *mix* were shown a picture of the prizes and they were told the market price of the *Lindt* chocolates in order to avoid them over- or underestimating the value of the non-monetary incentives (see Figure B1 in Appendix B for an exemplary screen). In addition, subjects were asked to indicate the attractiveness of *Lindt* chocolates and to rank the prizes according to their desirability in order to control for possibly different preferences (regarding chocolate) between individuals, especially because of stereotypes between men and women (Wiseman, 2010).

After the working period subjects were asked to self-assess their performance by selecting which performance ranking group (1-4) they believed they were in. In order to elicit subjects' accurate guesses about their own relative performance, we incentivized correct assumptions with 2 euros. Subjects then received feedback about their performance ranking, the number of mathematical problems they had solved and the proportion of correctly solved problems. At the end, participants had to answer a questionnaire on appreciation, motivation, other personal traits, and demographics. Finally, they received their payment. All treatments were randomly distributed across different sessions and times to ensure that treatment effects were not mixed up with, for example, general performance shocks, arising at different times of the day (Kube et al., 2012). For an overview of the experimental timeline, see Figure 3.2.

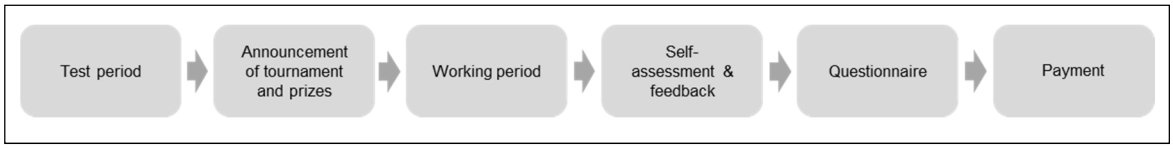


Figure 3.2: Experimental timeline

Our experiment took place from December 2016 to February 2017 at a large German university. 264 students[12] from various faculties − mainly economics (42%), engineering (24%), and industrial engineering (7%) − participated in our experiment, with 63% being male and an average age of 23 years (SD=4.0). According to a Kruskal-Wallis test, there were no differences according *age* (p=0.505) and *gender* (p=0.787) between treatments. The experiment lasted about 40 minutes and subjects received a fixed payment of 10 euros and an additional prize, depending on the treatment they were assigned to. Furthermore, subjects could earn an additional 2 euros if they self-assessed their performance correctly. We had a

[12] Two subjects had to be excluded as they filled in random numbers (one in *monetary* and one in *mix*).

total of 66 subjects in *control*, 65 in *monetary*, 66 in *nonmonetary*, and 65 in *mix*.

3.4 Results

3.4.1 The effectiveness of monetary, non-monetary, and mixed incentives

As outlined above, the aim of our experimental study was to analyze the effects of performance-related monetary, non-monetary, and mixed incentives on workers' performance. The number of correctly solved mathematical problems (*score*) served as a measure of performance and was considered our main outcome variable for the analyses. Figure B2 in Appendix B shows the mean of work performance in the different treatments. To ensure that possible differences in performance are not the result of ability, but of effort, we compared ability (measured by the time needed to solve the mathematical problems correctly in the test period (*testtime*) between treatments. According to a Kruskal-Wallis test, there were no differences in ability between treatments (p=0.310). Furthermore, the ability to solve simple mathematical problems did not significantly differ between men and women (Wilcoxon rank-sum test, p=0.132). Table 3.1 presents a statistical summary of the main variables included in the analyses, divided by treatments.

Variable	Description	Control (N=66)	Monetary (N=65)	Nonmonetary (N=66)	Mix (N=65)
Score	Number of correctly solved mathematical problems	93.88 (26.33)	101.65 (25.60)	101.64 (30.08)	102.83 (23.62)
Error	Number of incorrectly solved mathematical problems	5.08 (3.82)	5.68 (4.05)	5.58 (4.00)	5.55 (4.04)
Testtime	Time needed to solve five mathematical problems correctly	112.83 (71.60)	99.00 (58.15)	94.39 (54.86)	98.38 (67.63)
Age	Age in years	22.27 (3.36)	22.05 (2.46)	23.88 (5.93)	22.37 (3.16)
Gender	Dummy variable equals 1 if female	0.39	0.34	0.33	0.40

Note: Mean values are reported; standard deviations in parentheses.

Table 3.1: Descriptive statistics

Our results indicate a positive relationship between monetary incentives and performance: Performance was 8.28% higher in *monetary* than in *control* (Wilcoxon rank-sum test, p=0.175; t-test, p=0.089). In addition to monetary incentives, also non-monetary and mixed

incentives had a positive impact on performance (*nonmonetary* versus *control*: Wilcoxon rank-sum test, p=0.081; t-test, p=0.117; *mix* versus *control*: Wilcoxon rank-sum test, p=0.068; t-test, p=0.043). The average number of correctly solved mathematical problems was approximately 8.27% higher in *nonmonetary* than in *control*, and 9.53% higher in *mix* than in *control*. However, there were no significant differences in performance between *monetary* and *nonmonetary* (Wilcoxon rank-sum test, p=0.708; t-test, p=0.998). On average, subjects produced roughly the same number of correctly solved mathematical problems in these two treatments. Participants in pursuit of mixed incentives exerted slightly more effort than those in pursuit of pure monetary or pure non-monetary incentives, but this difference was not statistically significant (*mix* versus *monetary*: Wilcoxon rank-sum test, p=0.767; t-test, p=0.784; *mix* versus *nonmonetary*: Wilcoxon rank-sum test, p=0.961; t-test, p=0.801). Furthermore, when the subject pool is divided into winners (performance ranking groups 1 and 2) and losers (performance ranking groups 3 and 4), a Kruskal-Wallis test reveals no significant differences between treatments *monetary*, *nonmonetary*, and *mix* in these two clusters (winner: p=0.761; loser: p=0.849).

In addition, we analyzed work quality, measured by the number of incorrectly solved mathematical problems (*error*). According to our data, prizes have a slight negative impact on work quality in treatments *mix*, *monetary*, and *nonmonetary*. Nevertheless, according to a Kruskal-Wallis test, work quality did not differ significantly between treatments *monetary*, *nonmonetary*, *mix*, and *control* (p=0.840).

We further conducted OLS regressions to examine if the treatment effects found in the non-parametric/parametric tests were robust and to control for potential differences in abilities and other variables. Models (1)-(3) in Table 3.2 show the results for treatments *monetary*, *nonmonetary*, *mix*, and *control*, including all 262 subjects, while Models (4) and (5) address only treatments *monetary*, *nonmonetary*, and *mix*, and therefore contain 196 subjects. We included dummy variables for treatments *monetary*, *nonmonetary*, and *mix*, with *control* as reference group (in Model (4) *monetary* served as reference group, in Model (5) *mix* was the reference group). Furthermore, the variable *intrinsicmot* was included to control for intrinsic motivation. This was based on (dis)agreement with the statement "I had fun solving the task" measured on a seven-point Likert scale. The variable *belief* served as an indicator of which performance-ranking group subjects believed they were in, since a subject's payoff in a tournament does not depend only on his or her own performance, but also on that of the other

subjects. Finally, participants' personal characteristics such as *age*, *gender*, and ability (*testtime*), were also inserted in the model as control variables.

Score	**Model (1)** Reference: *control*	**Model (2)** Reference: *control*	**Model (3)** Reference: *control*	**Model (4)** Reference: *monetary*	**Model (5)** Reference: *mix*
Monetary (d)	7.767*	6.202	7.946**		-0.0426
	(1.71)	(1.43)	(2.01)		(-0.01)
Nonmonetary (d)	7.758	4.106	8.116**	0.240	0.197
	(1.58)	(0.84)	(2.02)	(0.05)	(0.05)
Mix (d)	8.952**	7.315*	7.747**	0.0426	
	(2.05)	(1.81)	(2.08)	(0.01)	
Gender (=1 if female)		-4.420	-1.438	-2.869	-2.869
		(-1.37)	(-0.48)	(-0.79)	(-0.79)
Age (years)		0.850*	0.667*	0.513	0.513
		(1.92)	(1.77)	(1.28)	(1.28)
Testtime (s.)		-0.109***	-0.0800***	-0.0903***	-0.0903***
		(-4.03)	(-3.76)	(-3.15)	(-3.15)
Intrinsicmot (1-7)			3.117***	2.545**	2.545**
			(3.22)	(2.28)	(2.28)
Belief (1-4)			-14.44***	-13.24***	-13.24***
			(-4.96)	(-3.23)	(-3.23)
Constant	93.88***	89.03***	105.2***	118.0***	118.1***
	(28.97)	(8.34)	(9.18)	(9.22)	(9.38)
R^2	0.018	0.099	0.288	0.237	0.237
adj. R^2	0.007	0.077	0.265	0.208	0.208
N	262	262	262	196	196

Note: Robust t-statistics in parentheses; significance levels are denoted as follows: * p < 0.10, ** p < 0.05, *** p < 0.01; (d) is for binary variable; the dependent variable is the number of correctly solved mathematical problems (*score*); Models (1)-(3) include all 262 subjects of all treatments, whereas Models (4) and (5) only include the 196 subjects of treatments *monetary*, *nonmonetary*, and *mix*.

Table 3.2: OLS regressions on work performance (*score*)

The results of our multivariate regressions, which were estimated with robust standard errors, confirm our previous findings. Model (3) shows a significant positive effect of monetary, non-monetary, and mixed incentives on performance, which is significant at the 5% level. Moreover, the positive effect of mixed prizes remains significant, whether or not explanatory variables are included (Models (1)-(3)). To conclude, these results support hypothesis 1.

Some of the other explanatory variables also show a significant effect on performance: First, the more intrinsically motivated subjects were to execute the task (*intrinsicmot*), the better their performance. There are no differences regarding the stated intrinsic motivation between the treatments (Kruskal-Wallis test, p=0.934), thus intrinsic motivation is not influenced by the nature of the incentive. Second, age had a significant positive impact on performance. Third, lower self-assessments of performance relative to the other participants (*belief*) correlated with lower actual performance. Finally, the lower the ability (*testtime*), the worse

the performance, although the coefficient is very small. Our main findings are summarized as follows:

Result 1. *Compared to the benchmark treatment control, monetary, non-monetary, and mixed incentives have a significant positive impact on performance.*

Model (4) in Table 3.2 was employed to shed light on whether monetary or non-monetary incentives are superior in terms of their effect on performance. The data revealed no significant difference in performance between the treatments *nonmonetary* and *monetary*. In addition, performance in *mix*, *monetary*, and *nonmonetary* did not significantly differ (see Model (5)). Hence, we note the following results:

Result 2. *There are no significant differences in the average incentive effects of treatments nonmonetary and monetary.*

Result 3. *Mixed incentives result in the highest performance, although there are no significant differences between treatments monetary, nonmonetary, and mix.*

3.4.2 Gender differences in the effectiveness of monetary, non-monetary, and mixed incentives

The initial results would seem to indicate that monetary, non-monetary and mixed incentives are equally suitable to enhance employees' performance. However, considering men and women separately reveals a more varied picture (see Figure 3.3).

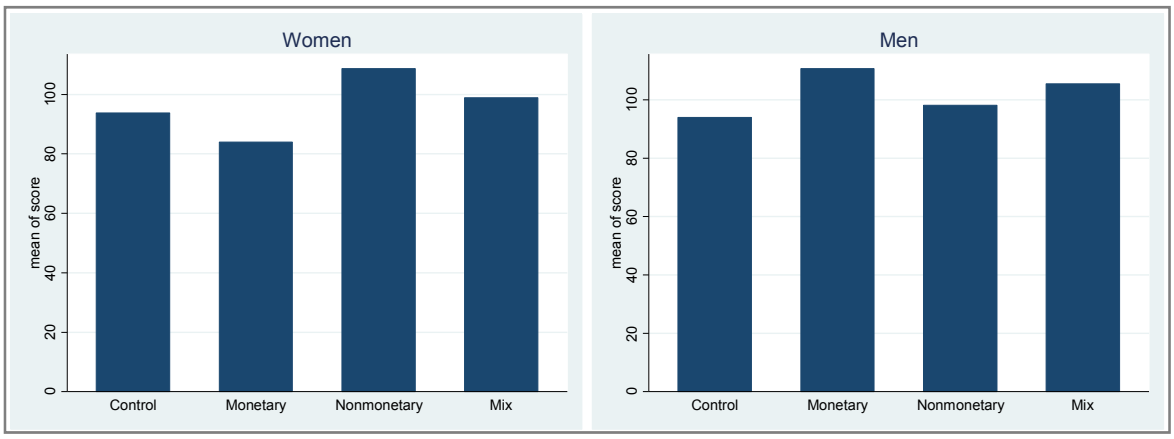


Figure 3.3: Gender differences in work performance over different treatments

In order to identify possible gender differences, we first conducted non-parametric tests (see Table 3.3). Monetary and mixed incentives had a significant positive impact on men's performance: Compared to the *control* group, the average number of mathematical problems was 17.83% higher in *monetary* and 12.28% higher in *mix*. Moreover, men's performance was 12.86% higher in pursuit of monetary incentives than in pursuit of non-monetary incentives, which was significant at the 10% level. In contrast, considering the female sample, monetary or mixed incentives had no significant impact on women's performance, whereas non-monetary incentives evoked a significant performance increase in the amount of 15.95%. Moreover, women's performance was 22.79% higher in the *nonmonetary* group than in the *monetary* group, significant at the 1% level. Also in the *mix* group women's performance was significantly higher compared with that in the *monetary* group.

		Monetary vs. control	**Nonmonetary vs. control**	**Mix vs. control**	**Monetary vs. nonmonetary**	**Monetary vs. mix**	**Nonmonetary vs. mix**
Men	Difference between scores in %	+17.83%***	+4.41%	+12.28%*	+12.86%*	+4.94%	-7.01%
	Wilcoxon rank-sum test (p-value)	0.005	0.322	0.069	0.063	0.378	0.327
Women	Difference between scores in %	-10.47%	+15.95%*	+5.42%	-22.79%***	-15.07%**	+9.99%
	Wilcoxon rank-sum test (p-value)	0.123	0.094	0.583	0.001	0.031	0.214

Note: ***, **, * indicate the statistical significance at the 1%, 5%, and 10% level, respectively; two-sided t-tests show similar results.

Table 3.3: Non-parametric tests − gender differences in performance

Furthermore, to analyze the gender differences in more detail, we conducted OLS regressions and inserted interaction terms of treatments and gender (*monetary* x *gender*, *nonmonetary* x *gender*, *mix* x *gender*) in our regression models (see Table 3.4).

First, we investigated the impact of monetary, non-monetary, and mixed incentives on men's performance. Compared to our benchmark treatment *control,* where no incentives were implemented, Model (2) in Table 3.4 reveals that monetary and mixed prizes had a highly significant positive effect on men's performance (*monetary*: p=0.000; *mix*: p=0.048). Although men's performance in *nonmonetary* was slightly higher than in *control*, this difference was not significant (p=0.413). Furthermore, Model (3) indicates that men's performance was significantly higher in pursuit of monetary incentives than in pursuit of non-monetary incentives, as predicted in hypothesis 4 (p=0.023). In Model (4), there was no

significant difference between men's performance in *mix* and *monetary*, or between their performance in *mix* and *nonmonetary* (*mix* versus *monetary*: p=0.126; *mix* versus *nonmonetary*: p=0.338); thus hypothesis 4 is only partly supported.

Score	**Model (1)** Reference: *control*	**Model (2)** Reference: *control*	**Model (3)** Reference: *monetary*	**Model (4)** Reference: *mix*
Monetary (d)	16.75***	16.62***		7.145
	(3.06)	(3.58)		(1.54)
Nonmonetary (d)	4.141	4.248	-12.17**	-5.027
	(0.65)	(0.82)	(-2.29)	(-0.96)
Mix (d)	11.54**	9.272**	-7.145	
	(2.06)	(1.99)	(-1.54)	
Monetary x gender	-26.56***	-24.76***		-20.85**
	(-3.04)	(-3.16)		(-2.57)
Nonmonetary x gender	10.82	12.66*	37.18***	16.33**
	(1.10)	(1.71)	(4.67)	(2.05)
Mix x gender	-6.460	-3.734	20.85**	
	(-0.72)	(-0.48)	(2.57)	
Gender (=1 if female)	-0.181	2.482	-22.20***	-1.348
	(-0.03)	(0.47)	(-3.76)	(-0.24)
Age (years)		0.682*	0.522	0.522
		(1.87)	(1.36)	(1.36)
Testtime (s.)		-0.0743***	-0.0827***	-0.0827***
		(-3.57)	(-2.94)	(-2.94)
Intrinsicmot (1-7)		3.057***	2.473**	2.473**
		(3.23)	(2.28)	(2.28)
Belief (1-4)		-14.77***	-13.78***	-13.78***
		(-5.21)	(-3.49)	(-3.49)
Constant	93.95***	103.7***	125.2***	118.0***
	(22.62)	(9.33)	(10.08)	(9.40)
R^2	0.087	0.345	0.311	0.311
adj. R^2	0.062	0.316	0.278	0.278
N	262	262	196	196

Note: Robust t-statistics in parentheses; significance levels are denoted as follows: * p < 0.10, ** p < 0.05, *** p < 0.01; (d) is for binary variable; the dependent variable is the number of correctly solved mathematical problems (*score*); Models (1) and (2) include all 262 subjects of all treatments, Models (3) and (4) only include the 196 subjects of treatments *monetary*, n*onmonetary*, and *mix*.

Table 3.4: OLS regressions on work performance (*score*) with regard to gender differences

These results are in line with the non-parametric tests. Therefore, one can conclude that, for men, pure monetary incentives are always better, or at least no worse, than any non-monetary or mixed incentives of equal market value. There were no differences in men's stated intrinsic motivation or in the belief about one's performance ranking between treatments (Kruskal-Wallis test, *intrinsicmot*: p=0.772; *belief*: p=0.699).

Result 4. *Men's performance in pursuit of monetary incentives is significantly higher than in pursuit of non-monetary incentives. There is no significant difference between men's performance in mix and monetary, or between their performance in mix and nonmonetary.*

In contrast, the pattern of women's performance in pursuit of monetary, non-monetary, and mixed incentives showed a completely different picture, as suggested by the non-parametric tests. In order to analyze the effect of treatments *monetary*, *nonmonetary*, and *mix* on women's performance, we conducted linear post estimation tests after the OLS regressions (e.g., H0: *monetary* + *monetary* x *gender* = 0) and report the relevant t-statistics and p-values in the following section. The results of Model (2) in Table 3.4 and linear post estimation tests show that whereas non-monetary incentives had a highly significant positive impact on women's performance (t(250)=3.13, p=0.002), monetary incentives had a negative impact on women's performance when compared to *control*, although this difference was not significant (t(250)=-1.29, p=0.198). There was no significant difference between women's performance in treatment *mix* and *control* (Model (2), t(250)=0.89, p=0.376). In addition, Model (3) shows that women's performance was significantly higher in pursuit of non-monetary than monetary incentives (t(186)=4.19, p=0.000). Furthermore, women performed significantly better in pursuit of non-monetary than mixed incentives (Model (4), t(186)=1.86, p=0.065), whereas their performance in pursuit of mixed incentives was significantly higher than in pursuit of monetary incentives (Model (4), t(186)=-2.05, p=0.042). Thus, our results are in line with hypothesis 5. These observed performance differences were not driven by intrinsic motivation, as a comparison of women's statement concerning fun at work did not show differences between treatments (Kruskal-Wallis test, p=0.613). Furthermore, there were no differences in their belief about one's performance ranking between treatments (Kruskal-Wallis test, p=0.350). The main findings are summarized in our fifth result:

Result 5. *Women's performance is significantly higher in pursuit of non-monetary incentives than in pursuit of monetary or mixed incentives. Furthermore, women's performance is significantly higher in pursuit of mixed than monetary incentives.*

Finally, we compared the performance of men and women in treatments *control*, *monetary*, *nonmonetary*, and *mix* (see Table 3.4). In pursuit of monetary incentives men's performance exceeded women's. This difference was highly significant (Model (3), p=0.000). Conversely, in the presence of non-monetary incentives, women had a significantly better performance than men (Model (3), p=0.000). However, in *mix* and *control*, there were no

significant differences in performance between men and women (*mix*: Model (4), p=0.807; *control*: Model (2), p= 0.640).

3.5 Discussion

In this section, we analyze subjects' motivation behind their effort decision and try to shed light on some possible explanations of our results, especially the impact of gender differences on the incentive effect of non-monetary and monetary prizes on performance.

When considering the whole subject pool without separating women from men, there was no significant difference in performance in pursuit of monetary and non-monetary incentives in our experimental data. This result contradicts both sides of the debate in the existing literature. While Hammermann and Mohnen (2014b) show that monetary incentives have a higher impact on performance than non-monetary incentives, Jeffrey (2009) concludes that non-monetary incentives outperform monetary incentives. These conflicting results may be driven by gender effects, as the proportion of men to women in these studies differs. Whereas in Hammermann and Mohnen's (2014b) study, the proportion of males was 64%, in Jeffrey's (2009) research, it was only 38%. These results are thus in line with our experimental results showing that men's performance is higher in pursuit of monetary incentives and women's performance is higher in pursuit of non-monetary incentives. Although gender differences are not discussed in Hammermann and Mohnen (2014b) and Jeffrey (2009), they might be a possible explanation for the mixed evidence in the literature pertaining to the effectiveness and superiority of monetary and non-monetary incentives.

Moreover, it might be argued that our results are triggered by the perceived attractiveness of the non-monetary incentive: *Lindt* chocolates may be less attractive for some subjects than, for example, massage vouchers used in Jeffrey (2009), and may therefore have a smaller incentive effect. The perceived attractiveness of the non-monetary incentive may further shape the gender differences in performance, as boxes of chocolates might be, in line with stereotypes, less attractive to men than women. Nevertheless, our examination does not support such arguments since only 21% of the subjects in *nonmonetary* and 14% of the

subjects in *mix* stated that *Lindt* chocolates are not attractive to them.[13] Moreover, there were no significant differences in the stated attractiveness of the prize between men and women in treatment *nonmonetary* (Wilcoxon rank-sum test, p=0.967) or in treatment *mix* (Wilcoxon rank-sum test, p=0.658). Unexpectedly, in *nonmonetary* only 18% of men indicated that chocolates are not attractive for them, compared to 27% of women. In *mix*, 13% of men and 15% of women stated no attractiveness. Therefore, we may rule out the possibility that gender differences in the effectiveness of non-monetary incentives on performance are the result of the prize's attractiveness.[14]

In addition, subjects were asked, ex post to the experiment, what impact the prizes had on their effort decision. Both men and women stated that monetary prizes had a significantly greater impact on performance than non-monetary prizes (Wilcoxon rank-sum test, women: p=0.039; men: p=0.000), despite the fact that women performed significantly better in pursuit of non-monetary incentives than in pursuit of monetary incentives. Moreover, 85% of subjects in *nonmonetary* (86% of women, 84% of men) and 95% in *monetary* (95% of women, 95% of men) stated their preference for cash over non-cash prizes.[15] Of the 65 participants in treatment *mix*, 71% (62% of women, 77% of men) stated that they preferred pure monetary prizes to a mix of monetary and non-monetary prizes, whereas 91% (81% of women, 97% of men) preferred mixed prizes instead of pure non-monetary prizes. These results confirm previous research which suggests that individuals state their preferences according to rational considerations, as money is the more rational choice, owing to its option value. Nevertheless, our results as well as other experimental studies show that often the

[13] Based on Hammermann and Mohnen (2014b) subjects were asked before the working period to rate the attractiveness of *Lindt* chocolates on a five-point Likert scale from 0, indicating no attractiveness, to 4, indicating full attractiveness. 21% (18% of men, 27% of women) stated no attractiveness (0), 15% (16% of men, 14% of women) rated *Lindt* chocolates as somewhat attractive (1), 36% (41% of men, 27% of women) as attractive (2 + 3), and 27% (25% of men, 32% of women) as very attractive (4). See Figure B3 in Appendix B for an overview of stated attractiveness of *Lindt* chocolates in *nonmonetary* and *mix*.

[14] Subjects' willingness to pay for *Lindt* chocolates was not retrieved, as willingness to pay is not a suitable measure for the perceived attractiveness of *Lindt* chocolates in our experimental setting. Although a person may only be willing to pay a low amount for *Lindt* chocolates due to, for example, budget constraints or justification concerns, he or she may rate *Lindt* chocolates as very attractive and thus aims to receive them.

[15] Participants in *nonmonetary* were asked at the end of the experiment if they would have preferred a monetary prize (with equal value) instead of the non-monetary prize, whereby agreement was measured on a seven-point Likert scale, from 1, indicating total disagreement, to 7, indicating total agreement. The same question was posed in reverse in *monetary*. Of the 66 participants in *nonmonetary*, 39 subjects strongly agreed (7) that they would have preferred a monetary prize, 17 subjects rated their response as agree (5+6), seven subjects were indifferent (4), and three subjects disagreed (1, 2, 3). Of the 65 subjects in *monetary*, 46 subjects strongly disagreed that they would have preferred a non-monetary prize (1), 16 disagreed (2+3), two subjects rated their response as neutral (4), and one agreed (6).

most preferred item is not actually the item which leads to the best performance (Hammermann & Mohnen, 2014b; Jeffrey, 2009; Kube et al., 2012; Shaffer & Arkes, 2009).

Another possible explanation of the observed gender differences in performance is the feeling of appreciation. According to Ellingsen and Johannesson (2007), appreciation and recognition are important drivers of employees' performance. Comparing statements of subjects' feelings of appreciation reveals that men felt much more appreciated by monetary prizes than by non-monetary prizes, and this difference is significant at the 5% level (Wilcoxon rank-sum test, p=0.041). Moreover, men stated a higher feeling of appreciation in *mix* than in *nonmonetary*, although this difference was not significant (Wilcoxon rank-sum test, p=0.526); further, there were no significant differences between *mix* and *monetary* (Wilcoxon rank-sum test, p= 0.142). In contrast, women felt significantly more appreciated by non-monetary prizes than by monetary prizes (Wilcoxon rank-sum test, p=0.018). Women's feeling of appreciation was also significantly higher in *mix* than in *monetary* (Wilcoxon rank-sum test, p=0.014). Furthermore, there were no significant differences in the women's stated feeling of appreciation between *mix* and *nonmonetary* (Wilcoxon rank-sum test, p=0.892). To conclude, the gender differences regarding the impact of monetary and non-monetary incentives on performance are reflected in the answers to the question about how appreciated subjects felt by their prizes (for an overview of the distributions of answers see Figure B4 in Appendix B). In addition to the feeling of appreciation, we asked subjects, ex post to the experiment, whether they were satisfied with their prize, as research has shown that satisfaction might have an influence on subjects' performance (Judge, Thoresen, Bono, & Patton, 2001). The answers are in line with those about the feeling of appreciation and with the performance pattern of men and women: whereas women stated a higher satisfaction with their prize in *nonmonetary* than in *monetary* (Wilcoxon rank-sum test, p=0.128), men stated that they were more satisfied with monetary than with non-monetary prizes, with a difference significant at the 5% level. However, our results about the feeling of appreciation and satisfaction can explain only part of the experimental results, as they do not explain the negative impact of monetary incentives on women's performance. While this may initially seem somewhat puzzling, existing research on competitions and performance pressure can help explain our results, as it shows that women reluctantly work in competitive environments (Niederle & Vesterlund, 2007) and falter under performance pressure (Azmat et al., 2016). By implementing a tournament, we created a competitive

environment. This competition and the related performance pressure may well have been intensified when monetary prizes were at stake: women stated that they felt significantly more performance pressure in *monetary* than in *control* (Wilcoxon rank-sum test, p=0.097).[16] This higher perceived pressure in *monetary* may have led to the observed negative effect on women's performance. In contrast, women's stated performance pressure in *nonmonetary* was lower than in *control*, although this difference was not statistically significant (Wilcoxon rank-sum test, p=0.163). Building on the findings of Heyman and Ariely (2004), we argue that non-monetary prizes may have reframed the competitive market as a more social market, thereby weakening the competitiveness of the tournament. Women may have therefore felt more comfortable to perform in pursuit of non-monetary incentives in a more social market, associated with lower competition, and thus exerted more effort than when pursuing monetary incentives. This, in turn, may have intensified the competition and the respective performance pressure in the tournament.

3.6 Conclusion

In a real-effort experiment, we analyzed the impact of performance-related non-monetary, monetary, and mixed incentives on employees' performance. Our data reveal three key findings. First, the experimental data suggest that monetary, non-monetary, and mixed incentives all have a significant positive impact on performance. Second, overall there are no significant differences between the treatments *monetary*, *nonmonetary*, and *mix*. Third, however, upon dividing the subject pool into men and women, we see a more differentiated picture: whereas men's performance is the highest in pursuit of monetary incentives, women's performance is higher in pursuit of non-monetary incentives than in pursuit of monetary or mixed incentives.

However, there are some limitations to the dataset and the experimental setting. The impact of monetary, non-monetary, and mixed incentives in our experimental setting was considered only within a short time period. Therefore, future research should analyze the effects of these incentives over a longer period to discover any long-term effects. In particular, it should be investigated whether the impact of non-monetary incentives

[16] Participants of all treatments were asked ex post, if they felt pressure to perform and to state the intensity of the pressure (from 1, indicating no pressure, to 7, indicating high pressure).

diminishes when the same incentives are used repeatedly. Company data or longer-term field studies would be suitable and useful for this purpose. Additionally, the composition of mixed incentives should be addressed in greater depth, as their impact may vary according to the proportion of monetary to non-monetary incentives.

Despite the constraints of the experimental setting, the study makes several contributions to the existing literature concerning monetary and non-monetary incentives. First, the results provide suggestive evidence that gender differences may clarify the mixed results regarding the impact of monetary and non-monetary incentives in the existing literature. Additionally, to the best of our knowledge, this is the first study to take into account gender differences when investigating the impact of monetary and non-monetary incentives on performance. Finally, we extend existing literature and provide evidence concerning the effectiveness of mixed incentives.

The comprehensive results of our experiment indicate that it is beneficial for companies to use non-monetary, monetary, and mixed incentives within competitive environments. Nevertheless, companies have to be aware that gender differences may play an important role regarding the effectiveness of these incentives. Understanding how these incentives enhance employee performance is crucial in implementing them effectively, since their underlying mechanisms may determine the amount of effort exerted by individuals in pursuit of a particular incentive. Employers can express their recognition and appreciation of employees' performance by means of incentives, however, employers have to be aware that monetary, non-monetary, and mixed incentives affect men and women and their feelings of acknowledgement differently. For instance, our findings suggest that men feel most valued when monetary rewards are given, while women feel much more appreciated by non-monetary incentives, which is reflected in employee performance. However, to generalize the results and to recommend an optimum incentive plan for companies – monetary, non-monetary, or mixed – future research should endeavor to obtain a deeper understanding of the motivational properties of non-monetary, monetary, and mixed incentives and their underlying psychological mechanisms, comprehensively and differentiating between genders.

4 The hidden value of non-monetary incentives: An experimental investigation on the effectiveness of non-monetary incentives and a cafeteria-style system[17]

4.1 Introduction

Many companies spend large amounts of money on non-monetary material incentives to motivate their employees. Whereas in 1996 only 26% of all US businesses used non-cash rewards, this ratio had increased to 84% by 2016. These figures show that non-monetary incentives have become increasingly popular over the past decades (Incentive Research Foundation, 2017). In addition to classical non-monetary incentives such as mobile phones or company cars, incentives such as complimentary beverages and lunches, laundry services, massages, childcare, or subsidized holidays are now widely used as well (Kauflin, 2017; Smith et al., 2015; Staufenbiel Institut, 2017; Staufenbiel Institut & Kienbaum, 2017; Zepelin, 2017). Although the use of non-monetary incentives is established in many companies, their effectiveness and underlying psychological mechanisms are rarely analyzed in the literature.

To date, there are only a few empirical studies that analyze the use of non-monetary benefits and their advantages to the company (Hammermann & Mohnen, 2014a; Oyer, 2008; Rajan & Wulf, 2006). Furthermore, existing research has analyzed the effectiveness of non-monetary incentives and widely demonstrated a positive effect of non-monetary incentives on performance (Bareket-Bojmel et al., 2017; Condly et al., 2003; Hammermann & Mohnen, 2014b; Jeffrey, 2009; Kube et al., 2012). However, to the best of our knowledge, extant research has not yet analyzed or compared the effectiveness of different non-monetary incentives of equivalent market value in one setting. Previous studies point out that non-monetary incentives might be not equally effective for all subjects, as each individual has different tastes and preferences (Jeffrey & Shaffer, 2007). To minimize this downside of non-monetary incentives, Jeffrey and Shaffer (2007) suggest using a cafeteria-style system,

[17] This chapter is based on a working paper co-authored by Alwine Mohnen. My contribution to the paper is summarized in Appendix D (signed by the authors in the examiners' copies of this dissertation). The working paper was presented at the *22nd Colloquium on Personnel Economics (COPE)*, March 2019 in Augsburg.

where employees can choose between different non-monetary incentives. Cafeteria-style systems are well established in companies, but their incentive impact is so far not examined. Hitherto, scholars have investigated the impact of having the choice between cash and non-cash incentives (Bareket-Bojmel et al., 2017; Kube et al., 2012; Shaffer & Arkes, 2009). For example, Shaffer and Arkes (2009) analyze preference reversals in the evaluation of cash versus non-cash incentives. In a side note, they further show that performance in the non-monetary incentive condition, where individuals had the choice of different prize options, does not differ compared to that in the monetary incentive condition. However, so far, there is no further empirical research on the impact of a cafeteria-style system, consisting of different non-monetary incentives, on performance.

Our experimental study aims (1) to investigate if different non-monetary incentives of equivalent value lead to the same performance level and (2) to examine the effectiveness of a cafeteria-style incentive system, under which subjects were given the choice between three different non-monetary incentives. Following Jeffrey (2009) and Hammermann and Mohnen (2014b), we define non-monetary incentives as tangible, non-cash incentives with market value. We conducted a laboratory experiment at a large German university with a real-effort task, where participants were randomly assigned to one of five treatment groups. Furthermore, we implemented a tournament, where participants could earn an additional non-monetary prize on top to their fixed wages according to their performance ranking. Whereas in the treatments *chocolate*, *cereals*, and *nuts*, the participants received predetermined non-monetary incentives (i.e., *Lindt* chocolates, *mymuesli* cereals, or *Seeberger* nuts, respectively), participants in the treatment *cafeteria* could choose between these three different non-monetary incentives and pick the one best fitting their individual needs. In our benchmark treatment *control*, subjects only received a fixed wage, without any additional incentives, but were informed about their ranking.

Our results suggest that non-monetary incentives of equivalent market value do not necessarily result in the same performance level. Not only does the market value of the non-monetary incentives matter, but their attractiveness also plays an important role for effectiveness. Furthermore, our data reveal that a cafeteria-style system with different non-monetary incentives has no significant positive impact on performance and that performance under a cafeteria-style incentive system is not significantly higher than that under predetermined incentives. In contrast, predetermined non-monetary incentives pushed the

top performers (i.e., the 35% best performing subjects in one experimental session) to a significantly higher performance level.

Hence, our study contributes in two ways to the literature. First, we analyze by means of a laboratory experiment the effect of different non-monetary incentives in one setting to determine if non-monetary incentives of equal market value result in the same performance level or if some other hidden values impact effectiveness. Second, our study contributes to the literature on non-monetary incentives by empirically analyzing the effects of a cafeteria-style incentive system on performance, where employees are offered the choice between different non-monetary incentives. We thus emphasize the importance of non-monetary incentives and point out the complexity of their underlying mechanisms and motivational properties.

The remainder of this chapter is structured as follows. Section 4.2 reviews the relevant literature and develops our research hypotheses, Section 4.3 presents the experimental design, followed by the presentation of the results and discussion in Section 4.4. Finally, Section 4.5 concludes with a short summary of our main results, limitations, and the management implications of our findings.

4.2 Literature overview and hypotheses

There are several empirical studies that have analyzed the use and associated advantages of non-monetary material incentives for companies. The main purposes providing non-monetary incentives seems to be alleviating cost efficiency concerns due to, for instance, scale economies or tax treatment (Hammermann & Mohnen, 2014a; Oyer, 2008), reducing employees' effort costs (Hammermann & Mohnen, 2014a; Oyer, 2008; Rajan & Wulf, 2006), and signaling good working conditions to attract target employees (Backes-Gellner & Tuor, 2010; Hammermann & Mohnen, 2014a). Furthermore, Hammermann and Mohnen (2014a) demonstrate that benefits are suitable means to reward good performance, show recognition to employees, and increase employees' work satisfaction.

In addition to the prevalence and use of non-monetary incentives across companies, the literature has further dealt with the impact of non-monetary and monetary incentives on performance in either gift-exchange games (Kube et al., 2012; Mahmood & Zaman, 2010)

or tournaments (Hammermann & Mohnen, 2014b; Jeffrey, 2009). Several studies, in different settings, show that non-monetary incentives have a positive impact on performance (Alonzo, 1996; Bareket-Bojmel et al., 2017; Condly et al., 2003; Jeffrey, 2009; Jeffrey & Adomdza, 2010; Kube et al., 2012; Sittenthaler & Mohnen, 2018). Existing studies further argue that non-monetary incentives might be seen as gifts and reflect acts of kindness. As a result, the time and effort invested by the employer into the gift are positively reciprocated by the employee by attaining higher performance (Baron & Kreps, 1999; Kube et al., 2012). By means of non-cash gifts, an employer can further signal that he or she knows the employee's preferences (Prendergast & Stole, 2001). According to Solnick and Hemenway (1996), tangible incentives have an emotional value and offer additional value besides the market value for recipients; it's the thought that counts and not only the incentive per se. Additionally, Jeffrey and Shaffer (2007) examine the motivational properties of non-monetary material incentives and introduce four psychological concepts: *justifiability*, *separability*, *evaluability*, and *social reinforcement*. The *justifiability* concept states that employees have to justify themselves when spending money on luxurious items. In contrast, if these items are bought by the employer and employees receive them for good performance, there is no need for the employees to justify consuming them. The *separability* argument is based on the mental accounting theory of Thaler (1999), stating that individuals have different mental accounts for different types of earnings. Jeffrey and Shaffer (2007) argue that subjects evaluate non-monetary incentives independently of their other income sources and, therefore, they might have a higher impact than monetary incentives. Monetary incentives are simply considered as "more salary" and thus might have a diminishing marginal utility. Regarding *evaluability*, non-monetary incentives allow subjects to mentally adjust the value of the benefit in two directions: downwards if the benefit seems to be out of reach and upwards if it is earnable. According to Jeffrey and Shaffer (2007), *social reinforcement* stems from the fact that non-cash incentives have a trophy value, as they are highly visible in the social environment. This visibility brings indirect attention to performance. As employees enjoy being recognized and acknowledged by their social environment for good performance, non-monetary incentives are worth receiving to obtain further recognition for their good performance. This is further underlined by Ellingsen and Johannesson (2007) who find that individuals have a desire for social esteem, they are concerned about what others think about them, and feel pride if they earn respect and attention from their social environment. Empirical evidence shows that status and social

recognition have a motivating power (Kosfeld & Neckermann, 2011; Neckermann et al., 2014; Neckermann & Frey, 2013). Non-monetary tangible incentives might be a suitable means of showing an employer's respect and an employee's status. Individuals will behave in a manner likely to attract social recognition (Bandura, 1986; Hubermann et al., 2004; Stajkovic & Luthans, 2003). Therefore, we can conclude that non-monetary incentives have a positive impact on performance, irrespective of their nature:

Hypothesis 1. *Non-monetary incentives have a positive impact on performance.*

Following rational considerations, non-monetary incentives of equivalent value should result in the same performance level. We therefore posit the following hypothesis:

Hypothesis 2a. *Non-monetary incentives of equal market value lead to the same performance level.*

However, the opposite could also hold true: the appreciation of non-monetary incentives may depend on individual preferences (Dzuranin et al., 2013; Hammermann & Mohnen, 2014b; Jeffrey & Shaffer, 2007). According to Jeffrey (2009), the effort decision in pursuit of non-monetary incentives is a function of the attractiveness of the respective non-monetary incentive. As not all non-monetary incentives are equally attractive, performance in pursuit of different non-monetary incentives might differ, although they have equal market values. The alternative hypothesis can thus be formulated as follows:

Hypothesis 2b. *There are differences in the impact of different non-monetary incentives of equal market value on performance.*

As preferences are individual to each employee, it is difficult to find one benefit appreciated by all employees and fitting all preferences (Jeffrey & Shaffer, 2007). According to Waldfogel (1993), gift givers are usually not perfectly informed about the recipient's preferences and, therefore, a non-monetary gift will create a deadweight loss; the greater the social distance between recipient and gift giver, the greater the deadweight loss. Therefore, a non-monetary incentive might be less motivating to the entire working population (Jeffrey & Shaffer, 2007). In contrast, a cafeteria-style incentive system, where employees can choose a benefit from a list or catalog, allows selecting benefits according to individual preferences (Dzuranin et al., 2013; Jeffrey & Shaffer, 2007). According to Jeffrey and Shaffer (2007), employers could thereby minimize the downside of non-monetary incentives

associated with different employees' preferences. In addition, there exists empirical evidence that the effort that individuals put in increases, if they are involved in the decision process of determining their compensation scheme (Kube et al., 2012). Mellizo, Carpenter, and Matthews (2014) further show, by means of a field study, that workers who have a say in how they would be compensated exert significantly more effort than those who have not a say in determining their payment schemes. Additionally, Charness, Cobo-Reyes, Jiménez, Lacomba, and Lagos (2012) analyze the impact of delegating wage decisions in a principal-agent relationship and show that the individuals, allowed to choose their own compensation, exhibited higher performance. This opportunity to decide might be perceived as an increase in self-determination and control. This may in turn enhance workers' performance. By implementing a cafeteria-style system, individuals are given the choice between different incentives and they can thus actively take part in determining their own compensation, which might have a positive impact on performance. Furthermore, Riener and Wagner (2019) analyze the effectiveness of non-monetary incentives on pupils' achievements in secondary schools and show that their self-reported learning effort increased, if they could choose one reward out of four (medal, homework voucher, parent letter, or surprise), for improving their performance. These insights from the literature support the following hypotheses:

Hypothesis 3. *A cafeteria-style incentive system, where subjects have the choice between different non-monetary incentives of equal market value, has a positive impact on performance.*

Hypothesis 4. *Performance is higher under a cafeteria-style incentive system where employees are offered the choice between different non-monetary incentives, than under predetermined non-monetary incentives.*

4.3 Experimental design

We conducted a real-effort experiment using the z-Tree software (Fischbacher, 2007) to analyze the effects of different non-monetary incentives and a cafeteria-style system on employee performance.[18] To this end students of a large German university were recruited by the online recruitment system ORSEE (Greiner, 2004) and were randomly assigned to one of five treatment groups. The experiment consisted of one working period lasting 15 minutes. Before the working period, participants had to go through a test period, to ensure they understood and became familiar with the task. The work task of our experiment was to solve simple mathematical problems. Each problem consisted of two equations and each equation comprised three one-digit numbers that had to be subtracted or added. Afterwards, the lower result of these equations had to be subtracted from the higher result and the final solution filled in an appropriate field. People were not allowed to use any devices such as calculators or pens and pencils. The task did not require prior knowledge, it was easy to understand, and work performance and quality were precisely measurable (Hammermann & Mohnen, 2014b). During the test period, participants were asked to solve five mathematical problems equal to the actual task and were only allowed to continue the experiment if all five mathematical problems were solved correctly. The time needed to solve the problems correctly (*testtime*) and the number of false answers (*testerror*) served as ability measures. After the test period, participants were told that their working period would be 15 minutes and that they could either solve mathematical problems or read articles from different fields (i.e., society, culture, travel, economics, science, and technology). During the working period, they could switch between these two options at any time. Reading articles thereby only served as an outside option and had no impact on their final payment. According to Corgnet et al. (2015), it is important to offer an outside option to avoid performance triggered by boredom and individuals only working because of the lack of desirable alternatives in a laboratory environment. The subjects were informed that their fixed wage was 10 euros and that they would have the opportunity to earn an additional prize that depended on their performance relative to the performances of the other participants. Each session included 22 participants. To avoid the creation of a middle group, we implemented four performance ranking groups: best group (ranking position 1), second group (ranking positions 2–8), third group (ranking positions 9–18), and worst group (ranking positions 19–

[18] This experiment is an extension of the experiment in *Chapter 3* of this thesis.

22) (Hammermann & Mohnen, 2014b; Jeffrey, 2009). The ranking positions were measured by the number of correctly solved mathematical problems. In the cases where two participants solved an equal number of mathematical problems, the proportion of correctly solved mathematical problems over all solved mathematical problems was decisive. If there was still a tie between two or more participants, the winner was decided by chance. Moreover, the value of prizes increased with the ranking group.

In our benchmark treatment *control* subjects only received a fixed wage without any additional prizes but were informed about their ranking position. We implemented three non-monetary treatments to analyze if there are differences in the incentive effects of non-monetary incentives of equal market value and to appeal to different tastes and preferences. In the treatment *chocolate*, the best subject received a large box of *Lindt* chocolates (*Lindt Pralinés Hochfein*) worth 10 euros, the subjects of the second group received a medium-sized box of *Lindt* chocolates (*Lindt Mini Pralinés* 100 g) worth 5 euros, the subjects of the third group a small box of *Lindt* chocolates (*Lindt Mini Pralinés* 44g) worth 2.50 euros, and the worst group received no prize. In the treatment *cereals*, participants could earn *mymuesli* cereals (1 box *mymuesli* + 2 *mymuesli2go*, 3 *mymuesli2go*, 1 *mymuesli2go Paleo*, or 0 respectively), and those in the treatment *nuts* received *Seeberger* nuts (6 different packages of *Seeberger* nuts, 4 different packages of *Seeberger* nuts, 2 different packages of *Seeberger* nuts, or 0, respectively). The prizes in the treatments *cereals* and *nuts* have an equivalent market value as those in the treatment *chocolate*. Moreover, we also implemented treatment *cafeteria*, where subjects could choose their prize out of *Lindt* chocolates, *mymuesli* cereals, or *Seeberger* nuts depending on their ranking group. This treatment allowed individuals to choose a prize, which fit their preferences best and therefore minimize the disadvantages of non-monetary incentives because of different employees' preferences (see Figure 4.1 for an overview of the prizes). *Lindt*, *mymuesli*, and *Seeberger* are highly priced products and popular brands in Germany (Lindt & Sprüngli, 2018; mymuesli, 2019; Seeberger, 2019).

Ranking / Treatment	Ranking position 1 (prize value: 10 €)	Ranking positions 2-8 (prize value: 5 €)	Ranking positions 9-18 (prize value: 2.50 €)	Ranking positions 19-22 (prize value: 0 €)
Chocolate	 *Lindt Pralinés Hochfein*	 *Lindt Mini Pralinés (100g)*	 *Lindt Mini Pralinés (44g)*	No prize
Cereals	 *1 box mymuesli + 2 mymuesli2go*	 *3 mymuesli2go*	 *1 mymuesli2go Paleo*	No prize
Nuts	 *6 packages of Seeberger nuts*	 *4 packages of Seeberger nuts*	 *2 packages of Seeberger nuts*	No prize
Cafeteria	 *Choice: Lindt, mymuesli or Seeberger*	 *Choice: Lindt, mymuesli or Seeberger*	 *Choice: Lindt, mymuesli or Seeberger*	No prize
Control	No prize	No prize	No prize	No prize

Figure 4.1: Overview of prizes

To avoid participants over- or underestimating the actual value and to ensure comparability, we revealed the price of the non-monetary incentives (Hammermann & Mohnen, 2014b). Before the working period, subjects in the treatments *chocolate*, *cereals*, *nuts*, and *cafeteria* were shown a picture of the prizes (see Figure C1 in Appendix C) and were asked to state the attractiveness of the respective prizes on a five-point Likert scale. Subjects in *cafeteria* had to rate the attractiveness of chocolates, cereals, and nuts. Furthermore, participants were asked to rank the prizes according to desirability. After the working period, the subjects were asked to self-assess their performance by choosing one of the four performance ranking groups in which they believed they belonged. Subjects were told that they could earn an additional 2 euros if their self-assessment was correct. After the self-assessment, subjects received feedback about the number of correctly solved mathematical problems, their performance ranking, and prize. In the treatment *cafeteria*, subjects were then asked to choose their prize. Finally, all treatments were complemented by a questionnaire on motivation, personal traits, and demographics, after which the subjects received their payment. For an overview of the experimental timeline, see Figure 4.2. The experimental sessions lasted around 40 minutes.

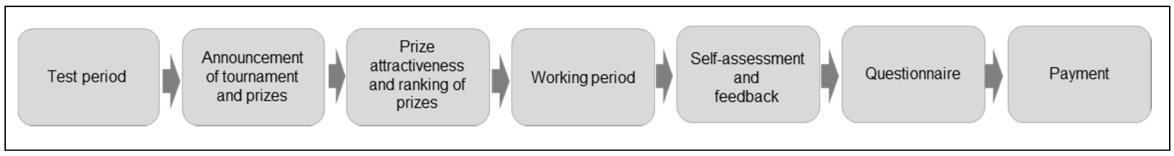


Figure 4.2: Experimental timeline

4.4 Results and discussion

4.4.1 Data set

From December 2016 to February 2017 a total of 330 individuals participated in our experiment and were randomly assigned to one of five treatment groups (*control*: 66; *chocolate*: 66; *cereals*: 65; *nuts*: 66; *cafeteria*: 66)[19]. Table 4.1 shows the demographics of our sample and the mean values of the main variables used for analyses.

Variable	Description	Control	Chocolate	Cereals	Nuts	Cafeteria
Subjects	Number of subjects	66	66	65	66	66
Gender	Female if gender=1	0.39	0.33	0.38	0.36	0.47
Age	In years	22.27 (3.36)	23.88 (5.93)	22.15 (3.05)	22.53 (4.00)	22.21 (3.05)
Testtime	Time subjects needed to correctly solve the mathematical problems in the test period (in sec.)	112.83 (71.60)	94.39 (54.86)	94.76 (51.51)	95.84 (53.68)	112.12 (71.98)
Testerror	Number of false answers in the test period	3.89 (8.80)	2.26 (6.62)	2.43 (5.83)	1.38 (3.35)	3.92 (10.37)
Score	Number of correctly solved mathematical problems	93.88 (26.33)	101.64 (30.08)	106.14 (28.67)	95.09 (33.73)	95.58 (28.41)
Worked	Number of mathematical problems worked	98.95 (26.74)	107.14 (30.96)	111.62 (29.37)	101.82 (34.94)	101.09 (29.46)

Note: Mean values are reported; standard deviations in parentheses.

Table 4.1: Descriptive statistics

The sample consisted of students of a large German university and their fields of study were mainly economics (40%), engineering (21%), and industrial engineering (11%). Overall, the average age of participants was 23 years (SD=4.05) and 39% were female. According to a Kruskal-Wallis test, there were no differences in age and gender between treatments (*age*: p=0.592; *gender*: p=0.583). Furthermore, there were no differences in ability, as measured

[19] One subject in *cereals* had to be excluded as he or she filled in random numbers.

by the time needed to correctly solve the five mathematical problems during the test period (*testtime*) and the number of false answers (*testerror*), between treatments (Kruskal-Wallis test, *testtime*: p=0.260; *testerror*: p=0.509).

4.4.2 Effectiveness of different non-monetary incentives

As outlined above, the aim of our study is to analyze the impact of different non-monetary incentives and a cafeteria-style incentive system on performance. Performance was measured as the number of correctly solved mathematical problems over the working period (*score*). On average, the subjects in the *control* treatment solved 93.88 (SD=26.33) mathematical problems correctly and 101.64 (SD=30.08) in the *chocolate* group. For the subjects in the treatments *cereals*, *nuts*, and *cafeteri*a, the respective performance levels were 106.14 (SD=28.67), 95.09 (SD=33.73), and 95.58 (SD=28.41). For a graphical overview of performance over the different treatments, see Figure 4.3.

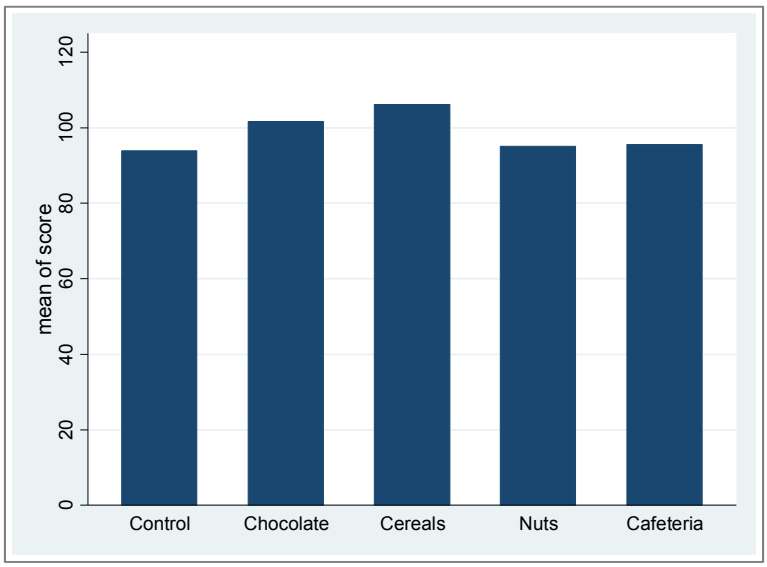


Figure 4.3: Work performance (*score*) over different treatments

As a first step, we conducted simple non-parametric tests to determine if scores were different between the benchmark treatment *control* and treatment groups *chocolate*, *cereals*, and *nuts* (see Table 4.2). Whereas performance in *chocolate* and *cereals* was significantly higher than in *control*, the performance difference between *nuts* and c*ontrol* did not reach statistical significance.

	Chocolate versus control			Cereals versus control			Nuts versus control		
	Overall	**Winners** (Ranking groups 1+2)	**Losers** (Ranking groups 3+4)	**Overall**	**Winners** (Ranking groups 1+2)	**Losers** (Ranking groups 3+4)	**Overall**	**Winners** (Ranking groups 1+2)	**Losers** (Ranking groups 3+4)
Difference between performance (compared to *control*)	8.27%*	9.20%**	7.47%*	13.06%**	11.37%***	14.04%***	1.29%	4.88%	-1.76%
Wilcoxon rank-sum test (p-value)	0.081	0.019	0.065	0.013	0.009	0.006	0.587	0.265	0.543

Note: Significance levels are denoted as follows: * p<0.10; ** p<0.05; *** p<0.01.

Table 4.2: Pairwise comparisons of mean performance (*score*) of the different treatments compared to the benchmark treatment *control*

To further analyze the effectiveness of different non-monetary incentives while controlling for other potential influences and analyze if our results are robust, we conducted ordinary least squares (OLS) regressions. The results of estimating the regressions with robust standard errors are reported in Table 4.3. Columns (1) and (2) show the results for the 263 subjects of treatments *control*, *chocolate*, *cereals*, and *nuts*, whereas column (3) is restricted to the winners (subjects in the first and second performance ranking groups) and column (4) to the losers (subjects in the third and fourth performance ranking groups) of these treatments. We inserted treatment dummies *chocolate*, *cereals*, *nuts*, and *cafeteria* as independent variables in our regression models, with *control* as the reference category. Moreover, we inserted the variable *testtime* (i.e., the time subjects needed to solve the mathematical problems correctly over the test period) to control for ability, as well as demographics (i.e., *gender*, *age*). Variable *intrinsicmot* was included to control for whether subjects have fun solving the task. Variable *belief* indicates the ranking group which subjects believed they belonged to, as the subjects' payoff depended on relative performance.

To analyze the impact of different non-monetary incentives on performance, we compared subjects' performance in the treatments *chocolate*, *cereals*, and *nuts* with the performance in the baseline treatment *control*. The results of the regressions support the descriptive statistics: Model (2) shows a significant positive effect of treatments *chocolate* and *cereals* on performance (*chocolate*: p=0.005; *cereals*: p=0.001), but no significant effect of *nuts* (p=0.211).

Score	**Model (1)**	**Model (2)**	**Model (3)** Winners (ranking groups 1+2)	**Model (4)** Losers (ranking groups 3+4)
Chocolate (d)	7.758	11.11***	11.19***	10.81***
	(1.58)	(2.85)	(2.77)	(2.65)
Cereals (d)	12.26**	12.34***	14.27***	11.61***
	(2.55)	(3.35)	(3.29)	(3.32)
Nuts (d)	1.212	4.855	6.818	2.847
	(0.23)	(1.26)	(1.52)	(0.70)
Gender (=1 if female)		5.362*	3.443	6.099**
		(1.86)	(1.00)	(2.11)
Age (in years)		0.165	0.910*	0.0162
		(0.50)	(1.93)	(0.05)
Testtime (in sec.)		-0.0664***	-0.0731***	0.00461
		(-3.13)	(-3.66)	(0.22)
Belief (1-4)		-20.74***	-8.294***	-15.70***
		(-7.91)	(-2.69)	(-5.28)
Inrinsicmot (1-7)		5.330***	2.712***	4.747***
		(5.26)	(2.64)	(4.06)
_cons	93.88***	115.1***	104.4***	91.80***
	(28.97)	(11.72)	(7.37)	(8.79)
R^2	0.027	0.470	0.319	0.453
adj. R^2	0.016	0.453	0.256	0.425
N	263	263	96	167
Posthoc tests (Wald test)				
cereals = nuts		p=0.055	p=0.101	p=0.033
cereals = chocolate		p=0.758	p=0.480	p=0.841
chocolate = nuts		p=0.129	p=0.313	p=0.086

Note: Robust t-statistics in parentheses; significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01; (d) is for binary variable; the dependent variable is the number of correctly solved mathematical problems (*score*); Models (1) and (2) include all 263 subjects of the treatments *control*, *chocolate*, *cereals*, and *nuts*, whereas Model (3) only includes the winners (96 subjects) and Model (4) the losers (167 subjects) of these treatments.

Table 4.3: OLS regressions on work performance

These effects remain stable when considering winners and losers separately (see Models (3) and (4)). Therefore, these results only partly support hypothesis H1. The results of Model (2) show that, on average, women's performance was significantly higher than men's (on average, 5.36 correctly solved mathematical problems). Moreover, the lower the ability of a participant, the lower was his or her performance, which was significant at the 1% significance level. Having fun solving mathematical problems (*intrinsicmot*) significantly increased productivity. Finally, the lower the participants' self-assessment of their own performance relative to other participants (*belief*), the worse their actual performance was. There were no differences in intrinsic motivation or in the self-assessment of their performance between treatments *chocolate*, *cereals*, *nuts*, and *control* (Kruskal-Wallis test, *intrinsicmot*: p=0.964; *belief*: p=0.315).

The main findings can be summarized as follows:

Result 1. *Compared to the benchmark treatment control, performance in the treatments chocolate and cereals is significantly higher. There are no significant differences in performance between treatments nuts and control.*

A second important question is whether chocolates, cereals, and nuts of equivalent value result in the same performance level. Experimental data show that, in terms of absolute values, *mymuesli* cereals resulted in the highest performance, followed by *Lindt* chocolates, and *Seeberger* nuts (see Table 4.1). Whereas the Wald test, conducted ex post to the OLS regression in Model (2) in Table 4.3, shows that the difference in performance between the treatment groups *cereals* and *nuts* was significant at the 10% level (in line with hypothesis H2b), performance in the *cereals* group and *chocolate* group did not significantly differ (in line with hypothesis H2a.). Although the performance level was higher in the *chocolate* group compared to the *nuts* group, the Wald test suggests that this difference missed the 10% significance level (in line with hypothesis H2a.). The literature suggests that the effectiveness of non-monetary incentives may depend on the attractiveness of prizes and on individual preferences (Hammermann & Mohnen, 2014b; Jeffrey, 2009; Jeffrey & Shaffer, 2007). According to Jeffrey (2009), performance is a function of the attractiveness of the benefit, which in turn depends on employees' valuation and may differ between them. To reveal the subjects' perceived attractiveness of the non-monetary incentives in our experiment, participants were asked before the working period to report how attractive they found the respective prizes and to rate them on a five-point Likert scale from 0 for no attractiveness to 4 for full attractiveness (cf. Hammermann & Mohnen, 2014b, see Table C1 in Appendix C). Figure 4.4 shows that the pattern of the stated attractiveness is in line with the performance patterns over the different treatments. Subjects stated a higher attractiveness for the prizes in *chocolate* and *cereals* compared to *nuts* (Wilcoxon rank-sum test, *chocolate* versus *nuts*: 2.20>1.76, p=0.086; *cereals* versus *nuts*: 2.31>1.76, p=0.036).

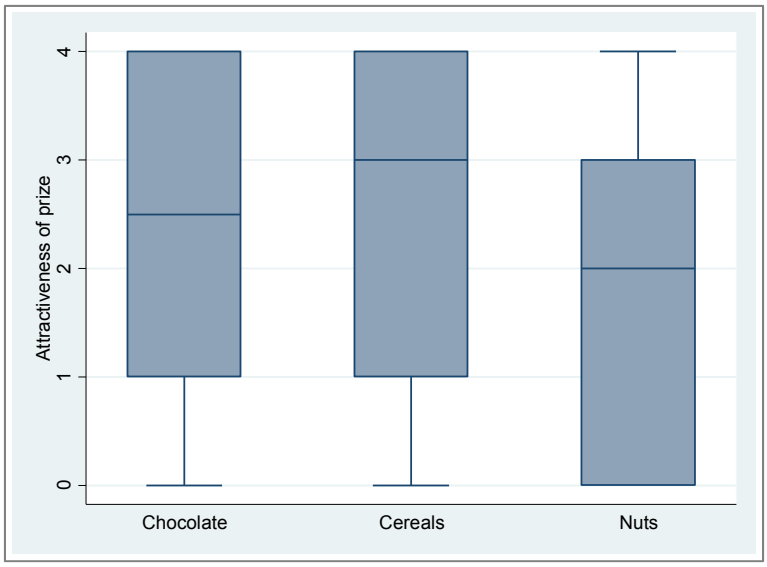


Figure 4.4: Distribution of stated attractiveness of prizes in the treatments *chocolate*, *cereals*, and *nuts*

Moreover, the Spearman rank correlation coefficient indicates a positive relationship between prize attractiveness and performance in the treatments *chocolate*, *cereals*, and *nuts*: the higher the stated prize attractiveness, the higher the performance ($\rho(197)=0.286$, $p=0.000$). We thus conclude that prize attractiveness plays an important role in the effectiveness of non-monetary incentives. Hence, we note the following:

Result 2: *Non-monetary incentives of equal market value do not inevitably lead to the same performance level. The attractiveness of prizes is a driving factor for the effectiveness of non-monetary incentives.*

We complemented the insights on work performance with evidence on work quality, measured as the ratio of correctly solved mathematical problems to the total number of problems worked (*quality*). Whereas our data reveal no significant difference in work quality between the treatments *chocolate* and *cereals*, and the benchmark treatment *control* (Wilcoxon rank-sum test, *chocolate* versus *control*: $p=0.888$; *cereals* versus *control*: $p=0.741$), work quality in the *nuts* group was significantly lower than in the *control* group (Wilcoxon rank-sum test, $p=0.076$). This negative work quality in the treatment *nuts* seems to be mainly driven by the winner cluster (*winner*: $p=0.065$; *loser*: $p=0.422$). To examine if the results of the non-parametric tests are robust, we conducted OLS regressions, which yielded similar results (see Table C2 in Appendix C). The lower work quality in the *nuts* group compared to the *control* group might be due to the unattractiveness of the incentive in

the *nuts* group: people stated a low attractiveness for *nuts* (i.e., the mean response was 1.76 on a Likert scale from 0 (*not attractive at all*) to 4 (*very attractive*)). Furthermore, the Spearman rank correlation coefficient indicates a positive relationship between prize attractiveness and work quality in the *nuts* treatment ($\rho(64)=0.224$; $p=0.075$); thus, the lower the prize attractiveness, the lower the work quality. This result agrees with the findings in the literature showing that people have individual views as to what extent rewards are desirable and appropriate. If they classify a reward as inequitable or inappropriate, this may lead to lower performance quality (Locke & Latham, 1990).

4.4.3 Effectiveness of a cafeteria-style incentive system

Our results so far indicate that the effectiveness of non-monetary incentives is – among other things – dependent on preferences. As it is difficult to find a benefit that fits all preferences and has an optimal incentive nature, we analyzed in a next step the effectiveness of a cafeteria-style incentive system, where individuals were offered the choice between three different non-monetary prizes. Therefore, they could choose the prize that best fit to their preferences. Unexpectedly, our data show that, compared to the benchmark treatment *control*, a cafeteria-style system had no significant impact on performance ($p=0.802$, see Table 4.4, Model (1)). Moreover, there were no differences in work quality compared with the *control* treatment group (Wilcoxon rank-sum test, $p=0.599$).

Result 3. *A cafeteria-system, where participants are offered the choice between different non-monetary incentives, has no significant positive impact on performance.*

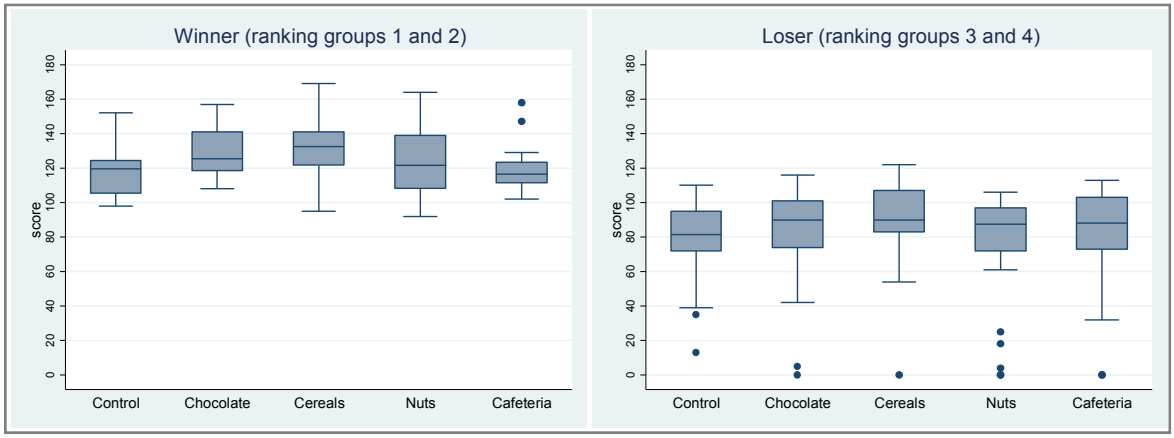
In a further step, we shed light on whether a cafeteria-style incentive system or preselected non-monetary incentives resulted in a higher productivity gain. Model (2) in Table 4.4 shows that the treatments *chocolate* and *cereals* resulted in a significantly higher output than the treatment *cafeteria* ($p=0.012$; $p=0.003$). Between the treatments *cafeteria* and *nuts*, there was no significant difference in performance ($p=0.281$). These results are therefore not in line with hypothesis H4, which states that a cafeteria-based system yields higher performance levels than preselected non-monetary incentives.

	Model (1) reference*: control*	**Model (2)** reference*: cafeteria*	**Model (3)** Winners reference*: cafeteria*	**Model (4)** Losers reference*: cafeteria*
Cafeteria (d)	0.943			
	(0.25)			
Chocolate (d)		10.50**	12.17***	8.061*
		(2.54)	(3.01)	(1.67)
Cereals (d)		11.71***	15.48***	9.162**
		(3.02)	(3.75)	(2.20)
Nuts (d)		4.426	8.046*	0.787
		(1.08)	(1.82)	(0.16)
Gender (=1 if female)	1.985	5.946**	2.388	6.052*
	(0.50)	(2.05)	(0.70)	(1.92)
Age (in years)	1.314**	0.174	1.175**	0.262
	(2.07)	(0.53)	(2.39)	(0.82)
Testtime (in sec.)	-0.0685***	-0.0750***	-0.0897***	-0.0143
	(-2.89)	(-3.08)	(-4.39)	(-0.54)
Belief (1-4)	-16.79***	-21.54***	-7.660**	-18.08***
	(-4.94)	(-7.00)	(-2.33)	(-5.18)
Intrinsicmot (1-7)	4.249***	4.440***	2.020*	3.718***
	(3.41)	(4.48)	(1.92)	(3.30)
_cons	87.65***	122.2***	101.5***	101.2***
	(4.96)	(12.21)	(7.57)	(8.81)
R^2	0.436	0.465	0.313	0.430
adj. R^2	0.409	0.448	0.249	0.401
N	132	263	96	167

Note: Robust t-statistics in parentheses; significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01; (d) is for binary variable; the dependent variable is the number of correctly solved mathematical problems (*score*); Models (1) includes 132 subjects of the treatments *control* and *cafeteria*, Model (2) includes 263 subjects of the treatments *chocolate*, *cereals*, *nuts*, and *cafeteria*, whereas Model (3) comprises only the winners (performance ranking groups 1+2) of these treatments, and Model (4) the losers (performance ranking groups 3+4) of these treatments.

Table 4.4: OLS regressions on work performance – consideration of the effectiveness of a cafeteria-style system

To analyze the differences in performance levels between the *cafeteria* treatment group and the treatment groups with predetermined prizes (*chocolate*, *cereals*, and *nuts*) in more detail, we divided the sample into winners (the first and second performance ranking groups) and losers (the third and fourth performance ranking groups). Our results show a similar picture (see Figure 4.5); nonetheless, the results of the separate consideration of winners and losers are worth mentioning.



Note: Boxplots display the distribution of performance of the winners (ranking groups 1 and 2) in the left graph and of the losers (ranking groups 3 and 4) in the right graph; dots represent outliers.

Figure 4.5: Distribution of performance of winners and losers over different treatments

Model (3) in Table 4.4 shows that, in the winner cluster, the incentive effect of the predetermined prizes compared to the cafeteria-style system seemed stronger. Performance in the treatments *chocolate*, *cereals*, and *nuts* was significantly higher than in the *cafeteria* treatment group (*chocolate* versus *cafeteria*: p=0.003; *cereals* versus *cafeteria*: p=0.000; *nuts* versus *cafeteria*: p=0.071). Considering the loser cluster, Model (4) reveals that performance in *chocolate* and *cereals* was significantly higher than in *cafeteria* (*chocolate* versus *cafeteria*: p=0.097; *cereals* versus *cafeteria*: p=0.030). By contrast, the average incentive effect in *nuts* and *cafeteria* did not significantly differ (p=0.869). Overall, one can conclude that it is more efficient to preselect non-monetary incentives at least not less efficient than using a cafeteria-style system, especially when considering top performers.

Result 4. *Performance in pursuit of a cafeteria-style system is not significantly higher than in pursuit of predetermined non-monetary incentives. By contrast, predetermined non-monetary incentives pushed top performers (i.e., subjects in the first and second ranking group) to a significantly higher performance level compared to the cafeteria-style system.*

These results are in contrast to Jeffrey and Shaffer (2007) who state that a cafeteria-style incentive system allows people to choose the benefits they like best, and should thereby minimize the downside of non-monetary incentives that every employee has different preferences, which makes it difficult to find an incentive fitting all employees. Our results are also in contrast to the literature showing that subjects' performance is positively affected if they are involved in the decision process of determining their compensation schemes

(Charness et al., 2012; Kube et al., 2012; Mellizo et al., 2014). Thus, the poor performance of a cafeteria-style system in our experimental setting might be surprising at a first glance. To find possible explanations for this result and analyze potential drivers of employees' effort decisions, we took a closer look at the questions asked before and after the experiment.

The feeling of appreciation might be a reason for the poor performance of a cafeteria-style system compared to predetermined non-monetary incentives. According to Ellingsen and Johannesson (2007), appreciation and recognition are important drivers of employees' performance. Further, Prendergast and Stole (2001) and Kube et al. (2012) state that the time an employer invests in searching for the gift matters. Non-monetary incentives are regarded as a kind act, often honored with positive reciprocity from employees. By preselecting non-monetary incentives, employers can further signal their knowledge of employees' preferences. In contrast, using a cafeteria system may signal that employers invested less time in searching for and finding the right non-monetary prizes for employees and do not know employees' preferences. This may be interpreted as a lower appreciation of good performance. Solnick and Hemenway (1996) show that it is the thought that counts and not only the incentive per se. To identify the feelings of appreciation, participants were asked to agree or disagree on a seven-point Likert scale with the statement "I felt appreciated by the prize". The mean responses to this question can be seen in Table C3 in Appendix C for every treatment, divided into the whole sample, winners, and losers. To analyze if subjects tended to feel less appreciated by the cafeteria-based incentive system than by preselected non-monetary benefits, we compared their statements about their feeling of appreciation over treatments (see Table 4.5).

Compared to subjects in the treatments *chocolate*, *cereals*, and *nuts*, subjects in *cafeteria* reported the lowest feelings of appreciation. Winners stated feeling a higher appreciation level in treatments with preselected benefits than in the cafeteria-based incentive system: Whereas winners in the *chocolate* treatment group stated a 61.84% higher feeling of appreciation than those in the *cafeteria* treatment group, winners' feelings of appreciation in the *cereals* and *nuts* treatment groups were 38.52% and 32.51% higher compared to the one in the *cafeteria* group. To conclude, the feeling of appreciation might be an explanation for the underperformance of the cafeteria-style system relative to the preselected non-monetary incentives.

	Chocolate versus cafeteria			Cereals versus cafeteria			Nuts versus cafeteria		
	Overall	Winner	Loser	Overall	Winner	Loser	Overall	Winner	Loser
Difference between stated feeling of appreciation (compared to *cafeteria*)	22.62%*	61.84%***	2.84%	18.69%*	38.52%**	8.52%	13.77%	32.51%	4.42%
Wilcoxon rank-sum test (p-value)	0.055	0.004	0.964	0.079	0.036	0.512	0.265	0.163	0.848

Note: Significance levels are denoted as follows: * p<0.10; ** p<0.05; *** p<0.01.

Table 4.5: Pairwise comparisons of mean feeling of appreciation compared to the treatment *cafeteria*

In addition to the feeling of appreciation, subjects were asked to state on a seven-point Likert scale what effect the possibility to choose a prize themselves (according to their ranking position) had on their effort decision. Of the 66 participants in the *cafeteria* group, 21 stated that it had no effect (1, 2, or 3), 12 were indifferent (4), and 33 stated that the possibility had a positive effect on their effort choice (5, 6, or 7). Although half of the subjects stated that the possibility to choose a prize had a positive effect on their performance, which is in line with the literature, we learned from other studies that subjects often state their opinions according to rational considerations and that statements and actual behavior may differ (Jeffrey, 2009; Kube et al., 2012; Shaffer & Arkes, 2009). This inconsistency is further shown in *Chapter 3* of this thesis; although women stated to prefer cash over non-cash incentives, their performance was higher under the non-cash condition.

Moreover, being involved in the decision process may also have a downside, as decision making can be costly (Shugan, 1980). Whereas in the treatments *chocolate*, *cereals*, and *nuts*, subjects were confined to accepting the prizes offered, subjects in *cafeteria* had to actively choose a prize, which namely offered the option to choose the prize best fitting their preferences, but might also have been associated with effort costs stemming from decision making.

According to the rational choice theory subjects have stable preferences and should thus choose the benefit which they had ranked highest (Neumann & Morgenstern, 1947). However, our data reveal that subjects' stated preferences did not always coincide with their actual choices: 67% of subjects in the *cafeteria* treatment group did not choose their most preferred prize, but a benefit that was lower in their stated performance ranking. A possible approach explaining this behavior might be justification concerns; for example, although subjects stated to prefer chocolates, they chose nuts or cereals, as they might have been

considered as the healthier choice. If they would have opted for chocolates, they might have had to justify their choice to themselves and/or their social surrounding. Closely linked to that is the concept of self-control, which refers to the ability to alter, restrain, or override one's responses and regulate behavior to align them with standards (e.g., social expectations or values). Responses that require self-control include, for example, overcoming unwanted impulses (e.g., resisting tempting sweets) (Baumeister, Vohs, & Tice, 2007; Tangney, Baumeister, & Luzio Boone, 2008). With regard to our study, having a high self-control would lead to overcoming the unwanted impulse of, for example, "choosing the unhealthy *Lindt* chocolates". Due to justification concerns and high self-control, subjects might thus have not chosen the benefit they valued most, but the one, for which they did not have to justify themselves. This might alternatively explain the lower effectiveness of the cafeteria-style system.

Scenario survey

In addition to the laboratory experiment, we conducted a scenario survey to obtain a deeper understanding of the poor performance of the cafeteria-style incentive system, in particular to analyze the justifiability concerns and the decision making behavior in the cafeteria-based system in more detail. Overall, 66 students[20] participated in this paper-pencil-study in November/December 2018, 43% of them being female. All participants were students at a large German university, mainly in economics (67%) and industrial engineering (14%), with an average age of 22 years (SD=2.62).

The students were asked to place themselves in the role of a temporary employee with the university library cataloging books in an electronic database for half a day together with nine other students. They were told that they would work on their own, independently of the other students, and each being placed in another room. The scenario assumes that the three employees who have correctly cataloged most books during the working time would receive an additional non-monetary bonus on top to their fixed wage. They would be offered the choice between the following three prizes: *Lindt* chocolates, *mymuesli* cereals, or *Seeberger* nuts (all of equal value). After reading the first part of the scenario, the students were asked to answer questions regarding motivation, preferences, and attractiveness of prizes. After

[20] One subject was excluded as he or she stated not being able to place him- or herself in the scenario.

that, the participants were presented the second part of the scenario: they were asked to imagine that they have now worked the whole morning cataloging books. When collecting their wage, the manager of the library told him or her that he or she is among the three best, and would thus receive a prize. The students were asked to state which prize they would choose and to answer questions related to appreciation, justifiability, self-control, decision making, and demographics.

To analyze justifiability concerns, following Jeffrey (2009), students were asked in the scenario if they would have to justify to themselves or their social surrounding when opting for chocolates, cereals, or nuts.[21] As a result, 32% of respondents stated to have to justify to themselves or their surrounding when choosing chocolates, and 18% when choosing cereals, or nuts. In contrast to the experimental results, stated preferences − regarding chocolates, cereals, and nuts − and stated choices are the same. Nevertheless, an OLS estimation shows that the cafeteria system loses its attractiveness if someone has to justify opting for chocolates (p=0.016, see Table C4 in Appendix C, Model (1)). Furthermore, the need for justifying choosing chocolates had a negative impact on motivation (p=0.034, Model (2)). We can thus conclude that justification concerns play an important role in the effectiveness of the cafeteria-style system.

Students were further asked questions regarding their decision making behavior, considering the cafeteria scenario. In general, individuals have different ways of making decisions: whereas some thoroughly evaluate all decision alternatives, others rely on an initial hunch (Hamilton, Shih, & Mohammed, 2016)[22]. Only 22% stated to have chosen the prize purely intuitively. Moreover, the survey data reveal that the more subjects thoroughly evaluate all decision alternatives and investigate all necessary information to decide which prize to choose, the lower is the stated motivation to work, significant at the 5% level (see Table C4 in Appendix C, Model (2)). Individuals who engage in lengthy deliberations might have

[21] Particularly, subjects were asked to state their agreement/disagreement with the following statements on a seven-point Likert scale: (1) It would be hard for me to justify to myself choosing *Lindt* chocolates and (2) It would be hard for me to justify to others choosing *Lindt* chocolates. The questions were also asked for *mymuesli* cereals and *Seeberger* nuts.

[22] Based on the *rational and intuitive decision style scale* of Hamilton et al. (2016), subjects were asked to state their dis-/agreement with six items regarding their decision behavior; see Figure C2 in Appendix C for an overview of the items.

higher effort costs than those who decide intuitively, which might in turn have a negative effect on motivation (Shugan, 1980).

To conclude, the survey data underline that justification concerns, as well as decision making behavior and the hereby related effort costs, have an impact on the effectiveness and success of the cafeteria-style system.

4.5 Conclusions, limitations, and implications

We have experimentally analyzed the impact of different non-monetary incentives of equal market value and a cafeteria-style incentive system on subjects' performance. Our study reveals three key findings: First, non-monetary incentives of equivalent market value do not necessarily result in the same performance level, as effectiveness further depends on the attractiveness of the benefit. Second, a cafeteria-style system, where subjects can choose between receiving different non-monetary incentives, has no positive impact on performance. Finally, performance in pursuit of a cafeteria-style system is not significantly higher than performance in pursuit of preselected non-monetary incentives. In contrast, predetermined non-monetary incentives pushed top performers to a significantly higher performance level. In addition to the feeling of less appreciation, justification concerns, and effort costs stemming from decision making may help explain the poor performance of the cafeteria-style system compared to predetermined prizes.

However, further work should be undertaken to analyze and better understand subjects' motives and their decision making process under a cafeteria-based incentive system. In particular, it would be useful to analyze in more detail subjects' effort costs stemming from the decision making of choosing a prize.

A further limitation of our study might be the nature of the non-monetary prizes and their small values, as different results might be obtained with high-priced luxury goods. Particularly with regard to a real working environment, where higher stakes are often involved, it would be worth changing the nature and implementing higher-priced benefits to see if the results remain robust. Although the ratio of non-monetary incentives and fixed payment is high in the experiment, absolute levels are low.

Another potential objection to the results is that we do not have depicted a long-term interaction and relationship between employer and employee in our experimental setting, which might have had an impact on the effectiveness of non-monetary incentives. According to Kube et al. (2012), there may be dynamic effects for long-term relationships, which cannot be found in one-shot interactions. Therefore, future research should address the long-term effects of non-monetary incentives. Furthermore, field data would be useful to check the robustness of our results in a real employer−employee relationship.

Despite these constraints, our experimental study allowed us to analyze and compare the effectiveness of different non-monetary incentives of equal value in one setting. Moreover, this study empirically analyzed the effectiveness of a cafeteria-style incentive system on performance, where subjects are offered the choice between different non-monetary incentives, in a laboratory environment. As such, the results underline the importance of non-monetary incentives and point out the complexity of their effects and underlying mechanisms.

Several important implications can be drawn from the experimental results. As the attractiveness of non-monetary incentives seems to play an important role in their effectiveness, it is thus important to reveal employees' preferences as to implement effective non-monetary incentives fitting individual preferences. Although it is difficult to find a benefit that fits all, a cafeteria-based incentive system seems not to solve this problem. In contrast, if employers aim to reward their employees for good performance by means of non-monetary incentives, it is beneficial for employers to preselect the benefits, especially when considering top performers. This result is contrary to expectations and seems counterintuitive at a first glance. However, by preselecting a benefit, employers can signal that they have invested time and effort in searching for and acquiring benefits that fit employees' preferences. Hereby, employers express appreciation to their employees. A predetermined incentive might be perceived rather as a reward for good performance, compared with an incentive chosen by oneself.

Non-monetary incentives thus have an impact, but they often affect behavior in unanticipated ways. Consequently, understanding the underlying (hidden) mechanisms of non-monetary incentives and how they increase performance is crucial to improve incentive systems and efficiently implement non-monetary incentives.

5 Conclusion

5.1 Summary of main research findings

The studies presented in this thesis analyzed the impact of variable payment cuts on performance and satisfaction (*Chapter 2*), the effectiveness of monetary, non-monetary, and mixed incentives (*Chapter 3*), and the effectiveness of different non-monetary incentives of equal value and a cafeteria-style incentive system (*Chapter 4*). In doing so, this research contributes to the incentive literature and emphasizes the importance of understanding human behavior in an employment relationship that cannot be fully explained by rational models. The key findings of chapters 2 through 4 are summarized as follows:

Chapter 2 investigated the impact of variable payment cuts on performance and satisfaction. Participants in a laboratory experiment were exposed to variable payment cuts of different levels to analyze how, and to what extent, they react to different pay cut levels. The results show that, in contrast to standard economic reasoning, variable payment cuts do not necessarily have a negative impact on performance. However, the level of the cut matters. Whereas a low cut and a medium cut both do not significantly affect performance, a high or total cut induces a significant decrease in performance equal to 22.8% and 41.2%, respectively. Moreover, the performance of subjects facing a high variable payment cut is reduced even more than what standard theory predicts. The data suggest that fairness concerns, demotivation, and goal setting aspects affect subjects' reaction to variable payment cuts. Furthermore, following a variable payment cut, subjects' satisfaction with the variable payment decreases significantly. Whereas a low cut leads to a 29.3% decrease in satisfaction, a medium cut decreases satisfaction by 58.1%. Following a high or total cut, subjects' stated satisfaction decreases on average by 64.4% and 67.1%, respectively. However, there is no significant difference in the decrease of satisfaction after a medium, high, or total cut. Thus, one can conclude that beyond a cut of certain value, satisfaction decreases to the same extent, irrespective of the level of the actual cut. Overall, this study points out that human behavior in reaction to variable payment cuts can strongly deviate from standard theoretical predictions; thus, it is essential to consider behavioral economic aspects when investigating the impact of variable payment cuts.

Chapter 3 analyzed the impact of monetary, non-monetary, and a combination of monetary and non-monetary incentives on employee performance. To this end, a laboratory experiment with a real-effort task was conducted. The data reveal that monetary, non-monetary, as well as a combination of monetary and non-monetary incentives, have a positive impact on performance. Overall, no significant differences in performance are found in response to monetary incentives, non-monetary incentives, or a combination of the two. However, considering them in terms of gender reveals quite a different picture. Men's performance in response to monetary incentives is significantly higher than that in response to non-monetary incentives. There is no significant difference between men's performance in pursuit of mixed and monetary incentives, or between their performance in pursuit of mixed and non-monetary incentives. In contrast, women's performance is significantly higher in response to non-monetary incentives than monetary or mixed incentives. The data indicate that these gender differences are not triggered by the perceived prize attractiveness, as men and women state that they consider the offered non-monetary prizes equally attractive. Possible explanations for the gender differences in the effectiveness of monetary and non-monetary incentives include revealed differences between men and women in the feeling of appreciation and perceived performance pressure in the tournament setting.

Chapter 4 described the first experimental study that analyzed the impact of different non-monetary incentives of equal market value in one setting. Furthermore, it analyzed the impact of a cafeteria-style incentive system, in which subjects could choose between different non-monetary incentives and, hence, according to their personal preferences. The results of this study indicate that non-monetary incentives of equal market value do not necessarily result in the same performance level, but the attractiveness of the prizes matters. Moreover, and surprisingly, the cafeteria-style incentive system has no significant impact on performance. In addition, performance in response to a cafeteria-style system is not significantly higher than in pursuit of preselected non-monetary incentives. In contrast, predetermined non-monetary incentives may result in a significantly higher performance level than a cafeteria-style system, especially among top performers. This result contradicts the suggestions of the existing literature (Dzuranin et al., 2013; Jeffrey & Shaffer, 2007) and seems counterintuitive at first glance because a cafeteria system allows the choice of non-monetary incentives in line with individual preferences. However, the data of the laboratory experiment and the additionally conducted scenario survey indicate that the feeling of less

appreciation, justification concerns, and effort costs related to decision making may help explain the poor performance of the cafeteria-style incentive system compared to one with predetermined prizes. The pre-selection of a single type of incentive by an employer indicates appreciation, as he or she invested time and effort in finding the right non-monetary incentive (Prendergast & Stole, 2001). This study points out the complexity of the (hidden) values and the underlying mechanisms of non-monetary incentives.

Overall, the results of this thesis suggest that standard economic models are not sufficient to explain the behavior of individuals and their performance in response to incentives. To analyze the effectiveness of incentives (and the impact of their cuts), various socio-psychological constructs and explanatory approaches need to be considered. Understanding how incentives enhance employee performance is crucial for implementing them effectively as the underlying mechanisms of incentives may determine the amount of effort exerted by individuals in response to a particular incentive.

5.2 Implications and further research

Objections have been raised regarding the external validity of laboratory experiments (Falk & Fehr, 2003; Falk & Heckmann, 2009). These criticisms suggest the limited generalizability of the results outlined in *Chapter 5.1*, especially with regard to the sample as well as the simplified depiction of circumstances. To validate and generalize the results of this thesis, future research should replicate the studies presented and use different methodological approaches, such as field studies or analyses based on company data. However, despite the constraints, the results of this thesis provide valuable insights and have several implications for both practice and future research.

Overall, this thesis indicates that subjects' behavior can deviate from standard theoretical predictions. Therefore, practitioners as well as scientists should take behavioral insights into account when deciding whether and how incentives should be implemented.

In contrast to rational economic theory, the results of this thesis show that people do not necessarily decrease their performance following a variable payment cut. Thus, if companies need to cut the variable payments of their employees (e.g., in times of economic crises), the level of the cut should be determined with care. Moreover, employers need to be aware that

variable payment cuts may cause demotivation and dissatisfaction, and these reactions are sensitive to the cut level. However, pay cuts may be perceived differently if the justifications for the cut are reasonable and credible for the workers (Chen & Horton, 2016; Greenberg, 1990; Lee & Rupp, 2007). Thus, future research should analyze whether and to what extent communicating the reasons for the variable payment cut has an impact on subjects' behavior.

Furthermore, the results of this thesis underline that, overall, monetary, non-monetary, and a combination of monetary and non-monetary incentives increase employee performance in a competitive environment. However, the understanding of how these incentives enhance performance is indispensable to implement them effectively. Not all kinds of incentives affect all individuals equally. In particular, this thesis shows that gender differences affect the effectiveness of these incentives. Furthermore, gender differences may help clarify the mixed results regarding the effectiveness of monetary and non-monetary incentives in the existing literature. By implementing incentives, employers need to be aware that incentives do not necessarily have the same impact on each employee; in fact, incentives might be perceived differently. Future research is needed to achieve a greater understanding of the most effective incentive for each type of employee. Besides gender differences, personal characteristics and cultural aspects may also play a role in the effectiveness of incentives and are worth further investigation.

Finally, the results of this thesis suggest that to effectively implement non-monetary tangible incentives, employers need to understand employees' preferences, as the attractiveness of incentives matters. Although finding a benefit that fits all is extremely challenging, a cafeteria-style incentive system, in which employees can choose a benefit themselves, does not solve the problem. In contrast, it is beneficial for the employer to preselect the benefits for the employees if they aim to reward them for good performance, especially when considering top performers. This result seems counterintuitive at first glance; however, the cafeteria-style system might diminish the reward and gift character of a non-monetary incentive, thus reducing its effectiveness. Moreover, by preselecting the non-monetary incentive, the employer can express a higher degree of appreciation to the employee. Future research, especially field studies or analyses based on corporate data, should focus on the effects of these incentives over a longer period to discover any long-term effects in real-life employer−employee relationships.

In general, this thesis adds valuable insights to the existing literature on incentives. To return to the quote of Prendergast, "*Incentives are the essence of economics*" (Prendergast, 1999, p. 7), stated at the beginning of my thesis, I can conclude that incentives are essential to motivate individuals; however, they often affect behavior in different and unexpected manners. Therefore, incentives must be used thoughtfully. When deciding whether and how incentives should be implemented, one needs to abandon the idea of a perfectly rational human being and take into account various socio-psychological aspects. Understanding the underlying psychological mechanisms of incentives and how they affect performance is crucial for implementing them effectively.

References

Adams, S. J. (1963). Toward an understanding of inequity. *Journal of Abnormal and Social Psychology, 67*(5), 422–436.

Aguinis, H., Joo, H., & Gottfredson, R. K. (2013). What monetary rewards can and cannot do: How to show employees the money. *Business Horizons, 56*(2), 241–249.

Akerlof, G. A. (1982). Labor contracts as partial gift exchange. *The Quarterly Journal of Economics, 97*(4), 543–569.

Akerlof, G. A., & Yellen, J. L. (1990). The fair wage-effort hypothesis and unemployment. *The Quarterly Journal of Economics, 105*(2), 255–283.

Alonzo, V. (1996). The trouble with money. *Incentive, 170*(2), 26–30.

Ariely, D., Gneezy, U., Loewenstein, G., & Mazar, N. (2009). Large stakes and big mistakes. *Review of Economic Studies, 76*(2), 451–469.

Ariely, D., Kamenica, E., & Prelec, D. (2008). Man's search for meaning: The case of Legos. *Journal of Economic Behavior & Organization, 67*(3-4), 671–677.

Ashraf, N., Bandiera, O., & Lee, S. S. (2014). Awards unbundled: Evidence from a natural field experiment. *Journal of Economic Behavior & Organization, 100*, 44–63.

Azmat, G., Calsamiglia, C., & Iriberri, N. (2016). Gender differences in response to big stakes. *Journal of the European Economic Association, 14*(6), 1372–1400.

Backes-Gellner, U., & Tuor, S. N. (2010). Avoiding labor shortages by employer signaling: On the importance of good work climate and labor relations. *Industrial and Labor Relations Review, 63*(2), 271–286.

Bandiera, O., Barankay, I., & Rasul, I. (2007). Incentives for managers and inequality among workers: Evidence from a firm-level experiment. *The Quarterly Journal of Economics, 122*(2), 729–773.

Bandura, A. (1986). *Social foundations of thought and action.* Englewood Cliffs: Prentice-Hall.

Bareket-Bojmel, L., Hochman, G., & Ariely, D. (2017). It's (not) all about the Jacksons: Testing different types of short-term bonuses in the field. *Journal of Management, 43*(2), 534–554.

Barnard, C. I. (1938). *The functions of the executive.* Cambridge: Harvard University Press.

Baron, J. N., & Kreps, D. M. (1999). *Strategic human resources: Frameworks for general managers.* New York: John Wiley & Sons, Inc.

Baumeister, R. F. (1984). Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology, 46*(3), 610–620.

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science, 16*(6), 351–355.

Besley, T., & Ghatak, M. (2008). Status incentives. *American Economic Review, 98*(2), 206–211.

Bewley, T. (2007). Fairness, reciprocity, and wage rigidity. In P. Diamond & H. Vartiainen (Eds.), *Behavioral economics and its applications* (pp. 157–188). Princeton and Oxford: Princeton University Press.

Blinder, A. S., & Choi, D. H. (1990). A shred of evidence on theories of wage stickiness. *The Quarterly Journal of Economics, 105*(4), 1003–1015.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review, 90*(1), 166–193.

Bonner, S. E., Hastie, R., Sprinkle, G. B., & Young, M. S. (2000). A review of the effects of financial incentives on performance in laboratory tasks: Implications for management accounting. *Journal of Management Accounting Research, 12*(1), 19–64.

Bonner, S. E., & Sprinkle, G. B. (2002). The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research. *Accounting, Organizations and Society, 27*(4-5), 303–345.

Bradler, C., Dur, R., Neckermann, S., & Non, A. (2016). Employee recognition and performance: A field experiment. *Management Science, 62*(11), 3085–3099.

Brandts, J., & Cooper, D. J. (2007). It's what you say, not what you pay: An experimental study of manager-employee relationships in overcoming coordination failure. *Journal of the European Economic Association, 5*(6), 1223–1268.

Brown, M. (2001). Unequal pay, unequal responses? Pay referents and their implications for pay level satisfaction. *Journal of Management Studies, 38*(6), 879–896.

Brüggen, A., & Strobel, M. (2007). Real effort versus chosen effort in experiments. *Economics Letters, 96*(2), 232–236.

Buser, T., Niederle, M., & Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics, 129*(3), 1409–1447.

Camerer, C., Babcock, L., Loewenstein, G., & Thaler, R. (1997). Labor supply of New York City cabdrivers: One day at a time. *The Quarterly Journal of Economics, 112*(2), 407–441.

Cameron, J., Banko, K. M., & Pierce, D. W. (2001). Pervasive negative effects of rewards on intrinsic motivation: The myth continues. *The Behavior Analyst, 24*(1), 1–44.

Cassar, L., & Meier, S. (2018). Nonmonetary incentives and the implications of work as a source of meaning. *Journal of Economic Perspectives, 32*(3), 215–238.

Chan, H. F., Frey, B. S., Gallus, J., & Torgler, B. (2014). Academic honors and performance. *Labour Economics, 31*, 188–204.

Chandler, D., & Kapelner, A. (2013). Breaking monotony with meaning: Motivation in crowdsourcing markets. *Journal of Economic Behavior & Organization, 90*, 123–133.

Charness, G. (2004). Attribution and reciprocity in an experimental labor market. *Journal of Labor Economics, 22*(3), 665–688.

Charness, G., Cobo-Reyes, R., Jiménez, N., Lacomba, J. A., & Lagos, F. (2012). The hidden advantage of delegation: Pareto improvements in a gift exchange game. *American Economic Review, 102*(5), 2358–2379.

Charness, G., Gneezy, U., & Henderson, A. (2018). Experimental methods: Measuring effort in economics experiments. *Journal of Economic Behavior & Organization, 149*, 74–87.

Charness, G., & Kuhn, P. (2007). Does pay inequality affect worker effort? Experimental evidence. *Journal of Labor Economics, 25*(4), 693–723.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics, 117*(3), 817–869.

Chen, D. L., & Horton, J. J. (2016). Research note−Are online labor markets spot markets for tasks? A field experiment on the behavioral response to wage cuts. *Information Systems Research, 27*(2), 403–423.

Choi, J., & Presslee, A. (2016). *The performance effects of tangible versus cash rewards: The mediating role of categorization.* AAA 2016 Management Accounting Section

(MAS) Meeting Paper. Retrieved June 10, 2019, from http://dx.doi.org/10.2139/ssrn.2645966.

Chou, Y. K. (2002). Testing alternative models of labour supply: Evidence from taxi drivers in Singapore. *The Singapore Economic Review, 47*(1), 17–47.

Clark, A. E. (1997). Job satisfaction and gender: Why are women so happy at work? *Labour Economics, 4*(4), 341–372.

Cohen, P. (2018, February 10). Where did your pay raise go? It may have become a bonus. *The New York Times.* Retrieved August 21, 2018, from https://www.nytimes.com/2018/02/10/business/economy/bonus-pay.html.

Cohn, A., Fehr, E., Herrmann, B., & Schneider, F. (2014). Social comparison and effort provision: Evidence from a field experiment. *Journal of the European Economic Association, 12*(4), 877–898.

Condly, S. J., Clark, R. E., & Stolovitch, H. D. (2003). The effects of incentives on workplace performance: A meta-analytic review of research studies. *Performance Improvement Quarterly, 16*(3), 46–63.

Cooper, D. J., Kagel, J. H., Lo, W., & Gu, Q. L. (1999). Gaming against managers in incentive systems: Experimental results with Chinese students and Chinese managers. *American Economic Review, 89*(4), 781–804.

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

Croson, R., & Gächter, S. (2010). The science of experimental economics. *Journal of Economic Behavior & Organization, 73*(1), 122–131.

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

Datta Gupta, N., Poulsen, A., & Villeval, M. C. (2013). Gender matching and competitiveness: experimental evidence. *Economic Inquiry, 51*(1), 816–835.

Deci, E. L. (1972). Intrinsic motivation, extrinsic reinforcement, and inequity. *Journal of Personality and Social Psychology, 22*(1), 113–120.

Deci, E. L. (1975). *Intrinsic motivation.* New York: Plenum Press.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*(6), 627–668.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior.* New York: Plenum Press.

Deci, E. L., & Ryan, R. M. (2002). *Handbook of self-determination research.* Rochester: The University of Rochester Press.

DellaVigna, S., List, J. A., Malmendier, U., & Rao, G. (2016). *Estimating social preferences and gift exchange at work.* NBR Working Paper No. 22043. Retrieved June 10, 2019, from http://www.nber.org/papers/w22043.

DellaVigna, S., & Pope, D. (2018). What motivates effort? Evidence and expert forecasts. *Review of Economic Studies, 85*(2), 1029–1069.

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

Dufwenberg, M., & Kirchsteiger, G. (2000). Reciprocity and wage undercutting. *European Economic Review, 44*(4-6), 1069–1078.

Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior, 47*(2), 268–298.

Dzuranin, A. C., Randolph, D., & Stuart, N. V. (2013). Doing more with less: Using noncash incentives to improve employee performance. *Journal of Corporate Accounting & Finance, 24*(5), 75–80.

Eckel, C. C., & Grossmann, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior, 23*(4), 281–295.

Elizur, D. (1994). Gender and work values: A comparative analysis. *The Journal of Social Psychology, 134*(2), 201–212.

Ellingsen, T., & Johannesson, M. (2007). Paying respect. *Journal of Economic Perspectives, 21*(4), 135–149.

Fahr, R., & Irlenbusch, B. (2000). Fairness as a constraint on trust in reciprocity: Earned property rights in a reciprocal exchange experiment. *Economics Letters, 66*(3), 275-282.

Falk, A., & Fehr, E. (2003). Why labour market experiments? *Labour Economics, 10*(4), 399–406.

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior, 54*(2), 293–315.

Falk, A., & Heckmann, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science, 326*(5952), 535–538.

Falk, A., & Ichino, A. (2006). Clean evidence on peer effects. *Journal of Labor Economics, 24*(1), 39–57.

Fang, M., & Gerhardt, B. (2012). Does pay for performance diminish intrinsic interest? *The International Journal of Human Resource Management, 23*(6), 1176–1196.

Fehr, E., & Falk, A. (2002). Psychological foundations of incentives. *European Economic Review, 46*(4-5), 687–724.

Fehr, E., & Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives, 14*(3), 159–181.

Fehr, E., Gächter, S., & Kirchsteiger, G. (1997). Reciprocity as a contract enforcement device: Experimental evidence. *Econometrica, 65*(4), 833–860.

Fehr, E., & Goette, L. (2007). Do workers work more if wages are high? Evidence from a randomized field experiment. *American Economic Review, 97*(1), 298–317.

Fehr, E., Kirchler, E., Weichbold, A., & Gächter, S. (1998). When social norms overpower competition: Gift exchange in experimental labor markets. *Journal of Labor Economics, 16*(2), 324–351.

Fehr, E., Kirchsteiger, G., & Riedl, A. (1993). Does fairness prevent market clearing? An experimental investigation. *The Quarterly Journal of Economics, 108*(2), 437–459.

Fehr, E., & List, J. A. (2004). The hidden costs and returns of incentives - trust and trustworthiness among CEOs. *Journal of the European Economic Association, 2*(5), 743–771.

Fehr, E., & Schmidt, K. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics, 114*(3), 817–868.

Fehr, E., & Schmidt, K. M. (2004). Fairness and incentives in a multi-task principal-agent model. *Scandinavian Journal of Economics, 106*(3), 453–474.

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

Frankfurter Allgemeine (2017, January 18). Vorstand der Deutschen Bank streicht seinen Bonus für 2016. *Frankfurter Allgemeine.* Retrieved June 14, 2019, from https://www.faz.net/aktuell/wirtschaft/unternehmen/vorstand-der-deutschen-bank-streicht-seinen-bonus-fuer-2016-14688681.html.

Freeman, R. B., & Kleiner, M. M. (2005). The last American shoe manufacturers: Decreasing productivity and increasing profits in the shift from piece rates to continuous flow production. *Industrial Relations, 44*(2), 307–330.

Frey, B. S. (1997). *Not just for the money: An economic theory of personal motivation.* Cheltenham: Edward Elgar Publishing.

Frey, B. S. (2007). Awards as compensation. *European Management Review, 4*(1), 6–14.

Frey, B. S. (2010). Geld oder Anerkennung? Zur Ökonomik der Auszeichnungen. *Perspektiven der Wirtschaftspolitik, 11*(1), 1–15.

Frey, B. S., & Jegen, R. (2001). Motivation crowding theory. *Journal of Economic Surveys, 15*(5), 589–611.

Friedman, D., & Cassar, A. (2004). *Economics lab: An intensive course in experimental economics.* New York: Routledge.

Friedman, D., & Sunder, S. (1994). *Experimental methods: A primer for economists.* Cambridge: Cambridge University Press.

Garbers, Y., & Konradt, U. (2014). The effect of financial incentives on performance: A quantitative review of individual and team-based financial incentives. *Journal of Occupational and Organizational Psychology, 87*(1), 102–137.

Gill, D., & Prowse, V. (2014). Gender differences and dynamics in competition: The role of luck. *Quantitative Economics, 5*(2), 351–376.

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives, 25*(4), 191–210.

Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics, 118*(3), 1049–1074.

Gneezy, U., & Rustichini, A. (2000). Pay enough or don't pay at all. *The Quarterly Journal of Economics, 115*(3), 791–810.

Gneezy, U., & Rustichini, A. (2004). Gender and competition at a young age. *American Economic Review, 94*(2), 377–381.

Goette, L., & Huffman, D. (2005). *Do emotions improve labor market outcomes?* IZA Discussion Papers 1895. Retrieved June 11, 2019, from http://hdl.handle.net/10419/33147.

Goette, L., Huffman, D., & Fehr, E. (2004). Loss aversion and labor supply. *Journal of the European Economic Association, 2*(2-3), 216–228.

Gonzalez, C., Best, B., Healy, A. F., Kole, J. A., & Bourne, L. E. (2011). A cognitive modeling account of simultaneous learning and fatigue effects. *Cognitive Systems Research, 12*(1), 19–32.

Grant, A. M., & Gino, F. (2010). A little thanks goes a long way: Explaining why gratitude expressions motivate prosocial behavior. *Journal of Personality and Social Psychology, 98*(6), 946–955.

Greenberg, J. (1990). Employee theft as a reaction to underpayment inequity: The hidden cost of pay cuts. *Journal of Applied Psychology, 75*(5), 561–568.

Greiner, B. (2004). An online recruitment system for economic experiments. In K. Kremer & V. Macho (Eds.), *Forschung und wissenschaftliches Rechnen. GWDG-Bericht Nr. 63* (pp. 79–93). Göttingen: Gesellschaft für Wissenschaftliche Datenverarbeitung.

Grossmann, S. J., & Hart, O. D. (1983). An analysis of the principal-agent problem. *Econometrica, 51*(1), 7–45.

Hamilton, K., Shih, S.-l., & Mohammed, S. (2016). The development and validation of the rational and intuitive decision styles scale. *Journal of Personality Assessment, 98*(5), 523–535.

Hammermann, A., & Mohnen, A. (2014a). Who benefits from benefits? Empirical research on tangible incentives. *Review of Managerial Science, 8*(3), 327–350.

Hammermann, A., & Mohnen, A. (2014b). The pric(z)e of hard work: Different incentive effects of non-monetary and monetary prizes. *Journal of Economic Psychology, 43*, 1–15.

Hansen, L. W., & Weisbrod, B. (1972). Toward a general theory of awards, or, do economists need a hall of fame? *Journal of Political Economy, 80*(2), 422–431.

Healy, A. F., Kole, J. A., Buck-Gengler, C. J., & Bourne, L. E. (2004). Effects of prolonged work on data entry speed and accuracy. *Journal of Experimental Psychology: Applied, 10*(3), 188–199.

Helmore, E. (2016, March 08). Wall Street bonuses fell 9% in 2015 to average of $146,200 after profits decline. *The Guardian.* Retrieved January 16, 2019, from https://www.theguardian.com/business/2016/mar/08/wall-street-bonuses-fell-2015-new-regulations-decline-profits.

Hennig-Schmidt, H., Sadrieh, A., & Rockenbach, B. (2010). In search of workers' real effort reciprocity−A field and a laboratory experiment. *Journal of the European Economic Association, 8*(4), 817–837.

Heyman, J., & Ariely, D. (2004). Effort for payment: A tale of two markets. *Psychological Science, 15*(11), 787–793.

Holmstrom, B. (1979). Moral hazard and observability. *The Bell Journal of Economics, 10*(1), 74–91.

Holmstrom, B. (1989). Agency costs and innovation. *Journal of Economic Behavior & Organization, 12*(3), 305–327.

Hubermann, B. A., Loch, C. H., & Önçüler, A. (2004). Status as a valued resource. *Social Psychology Quarterly, 67*(1), 103–114.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*(2), 139–155.

Incentive Research Foundation (2017). *IRF 2017 trends study.* Retrieved July 17, 2018, from http://theirf.org/research/irf-2017-trends-study/1940/.

Jalava, N., Schrøter Joensen, J., & Pellas, E. (2015). Grades and rank: Impact of non-financial incentives on test performance. *Journal of Economic Behavior & Organization, 115*, 161–196.

Jeffrey, S. A. (2009). Justifiability and the motivational power of tangible noncash incentives. *Human Performance, 22*(2), 143–155.

Jeffrey, S. A., & Adomdza, G. K. (2010). Incentive salience and improved performance. *Human Performance, 24*(1), 47–59.

Jeffrey, S. A., & Shaffer, V. (2007). The motivational properties of tangible incentives. *Compensation & Benefits Review, 39*(3), 44–50.

Jenkins, D. G., Mitra, A., Gupta, N., & Shaw, J. D. (1998). Are financial incentives related to performance? A meta-analytic review of empirical research. *Journal of Applied Psychology, 83*(5), 777–787.

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics, 3*(4), 305–360.

Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The job satisfaction−Job performance relationship: A qualitative and quantitative review. *Psychological Bulletin, 127*(3), 376–407.

Kahneman, D., Knetsch, J. L., & Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *American Economic Review, 76*(4), 728–741.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263–292.

Kamenica, E. (2012). Behavioral economics and psychology of incentives. *Annual Review of Economics, 4*, 427–452.

Kauflin, J. (2017, November 14). The 20 best company benefits and 11 companies that offer them. *Forbes.* Retrieved March 22, 2019, from https://www.forbes.com/sites/jeffkauflin/2017/11/14/the-20-best-company-benefits-and-11-companies-that-offer-them/#5b03351947f1.

Kawaguchi, D., & Ohtake, F. (2007). Testing the moral theory of nominal wage rigidity. *Industrial and Labor Relations Review, 61*(1), 59–74.

Kelly, K., Presslee, A., & Webb, R. A. (2017). The effects of tangible rewards versus cash rewards in consecutive sales tournaments: A field experiment. *The Accounting Review, 92*(6), 165–185.

Knez, M., & Simester, D. (2001). Firm-wide incentives and mutual monitoring at Continental Airlines. *Journal of Labor Economics, 19*(4), 743–772.

Kosfeld, M., & Neckermann, S. (2011). Getting more work for nothing? Symbolic awards and worker performance. *American Economic Journal: Microeconomics, 3*(3), 86–99.

Kosfeld, M., Neckermann, S., & Yang, X. (2017). The effects of financial and recognition incentives across work contexts: The role of meaning. *Economic Inquiry, 55*(1), 237–247.

Kräkel, M. (2007). *Organisation und Management* (3rd ed.). Tübingen: Mohr Siebeck.

Kube, S., Maréchal, M. A., & Puppe, C. (2012). The currency of reciprocity: Gift exchange in the workplace. *American Economic Review, 102*(4), 1644–1662.

Kube, S., Maréchal, M. A., & Puppe, C. (2013). Do wage cuts damage work morale? Evidence from a natural field experiment. *Journal of the European Economic Association, 11*(4), 853–870.

Kühl, S. (2009). Experiment. In S. Kühl, P. Strodtholz, & A. Taffertshofer (Eds.), *Handbuch Methoden der Organisationsforschung. Quantitative und Qualitative Methoden* (pp. 534–557). Wiesbaden: VS Verlag für Sozialwissenschaften.

Lacetera, N., & Macis, M. (2010). Do all material incentives for pro-social activities backfire? The response to cash and non-cash incentives for blood donations. *Journal of Economic Psychology, 31*(4), 738–748.

Laffont, J.-J., & Martimort, D. (2002). *The theory of incentives: The principal-agent model.* Princeton: Princeton University Press.

Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review, 90*(5), 1346–1361.

Lee, D., & Rupp, N. G. (2007). Retracting a gift: How does employee effort respond to wage reductions? *Journal of Labor Economics, 25*(4), 725–761.

Levitt, S. D., List, J. A., Neckermann, S., & Sadoff, S. (2016). The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. *American Economic Journal: Economic Policy, 8*(4), 183–219.

Lindt & Sprüngli (2018). *Annual Report 2018.* Retrieved June 12, 2019, from https://report.lindt-spruengli.com/18/ar/app/themes/lindtspruengli/dist/pdf/WEB_GB18_Gesamt_en.pdf.

List, J. A., & Shogren, J. F. (1998). The deadweight loss of Christmas: Comment. *American Economic Review, 88*(5), 1350–1355.

Locke, E. A., Bryan, J. F., & Kendall, L. M. (1968). Goals and intentions as mediators of the effects of monetary incentives on behavior. *Journal of Applied Psychology, 52*(2), 104–121.

Locke, E. A., & Latham, G. P. (1990). Work motivation and satisfaction: Light at the end of the tunnel. *Psychological Science, 1*(4), 240–246.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist, 57*(9), 705–717.

Locke, E. A., & Latham, G. P. (2013). *New developments in goal setting and task performance* (1 ed). New York: Routledge.

Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal setting and task performance: 1969-1980. *Psychological Bulletin, 90*(1), 125–152.

Loewenstein, G. (1999). Experimental Economics from the vantage−Point of Behavioral Economics. *The Economic Journal, 109*(453), 25-34.

Luthans, K. (2000). Recognition: A powerful, but often overlooked, leadership tool to improve employee performance. *The Journal of Leadership Studies, 7*(1), 31-39.

Mahmood, S., & Zaman, A. (2010). Monetary and non-monetary gift exchange. *The Pakistan Development Review, 49*(4), 719–740.

Malmendier, U., & Tate, G. (2009). Superstar CEOs. *The Quarterly Journal of Economics, 124*(4), 1593–1638.

Markham, S. E., Scott, D. K., & McKee, G. H. (2002). Recognizing good attendance: A longitudinal, quasi-experimental field study. *Personnel Psychology, 55*(3), 639–660.

Mas, A. (2006). Pay, reference points, and police performance. *The Quarterly Journal of Economics, 71*(3), 783–821.

Mas, A., & Pallais, A. (2017). Valuing alternative work arrangements. *American Economic Review, 107*(12), 3722–3759.

Masclet, D., Peterle, E., & Larribeau, S. (2015). Gender differences in tournament and flat-wage schemes: An experimental study. *Journal of Economic Psychology, 47*, 103–115.

McGregor, J. (2018, January 18). Why many companies are giving bonuses − not raises − after the new tax cuts. *The Washington Post.* Retrieved August 21, 2018, from https://www.washingtonpost.com/news/on-leadership/wp/2018/01/18/why-many-companies-are-giving-bonuses-not-raises-after-the-new-tax-cuts/?utm_term=.a09899caeaa.

Mellizo, P., Carpenter, J., & Matthews, P. H. (2014). Workplace democracy in the lab. *Industrial Relations Journal, 45*(4), 313–328.

Mellström, C., & Johannesson, M. (2008). Crowding out in blood donation: Was Titmuss right? *Journal of the European Economic Association, 6*(4), 845–863.

Moyer, L. (2016, March 07). Wall Street bonuses fell in 2015, and 2016 isn't looking rosy. *The New York Times.* Retrieved January 16, 2019, from https://www.nytimes.com/2016/03/08/business/dealbook/wall-street-bonuses-fell-in-2015.html.

mymuesli (2019). *mymuesli im Markencheck.* Retrieved June 12, 2019, from https://www.mymuesli.com/presse/wp-content/uploads/2019/02/LZ_25.01.19_Innofact.pdf.

Neckermann, S., Cueni, R., & Frey, B. S. (2014). Awards at work. *Labour Economics, 31*, 205–217.

Neckermann, S., & Frey, B. S. (2013). And the winner is…? The motivating power of employee awards. *The Journal of Socio-Economics, 46*, 66–77.

Neumann, J. von, & Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd ed.). Princeton: Princeton University Press.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics, 122*(3), 1067–1101.

Oyer, P. E. (2008). Salary or benefits? *Research in Labor Economics, 28*, 429–467.

Paarsch, H. J., & Shearer, B. (2000). Piece rates, fixed wages, and incentive effects: Statistical evidence from payroll records. *International Economic Review, 41*(1), 59–92.

Pereira, P. T., Silva, N., & Silva, J. A. e. (2006). Positive and negative reciprocity in the labor market. *Journal of Economic Behavior & Organization, 59*(3), 406–422.

Pokorny, K. (2008). Pay−but do not pay too much: An experimental study on the impact of incentives. *Journal of Economic Behavior & Organization, 66*(2), 251–264.

Pouliakas, K. (2010). Pay enough, don't pay too much or don't pay at all? The impact of bonus intensity on job satisfaction. *KYKLOS, 63*(4), 597–626.

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

Prendergast, C., & Stole, L. (2001). The non-monetary nature of gifts. *European Economic Review, 45*(10), 1793–1810.

Presslee, A., Vance, T. W., & Webb, R. A. (2013). The effects of reward type on employee goal setting, goal commitment, and performance. *The Accounting Review, 88*(5), 1805–1831.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review, 83*(5), 1281–1302.

Rajan, R. G., & Wulf, J. (2006). Are perks purely managerial excess? *Journal of Financial Economics, 79*(1), 1–33.

Riener, G., & Wagner, V. (2019). On the design of non-monetary incentives in schools. *Education Economics, 27*(3), 223–240.

Ross, S. A. (1973). The economic theory of agency: The principal's problem. *American Economic Review, 63*(2), 134–139.

Samuelson, P. A., & Nordhaus, W. D. (1992). *Principles of economics* (14th ed.). New York: McGraw-Hill.

Seeberger (2019). *Marke des Jahrhunderts.* Retrieved June 12, 2019, from https://www.seeberger.de/genuss-magazin/artikeldetailseite/marke-des-jahrhunderts.

Shaffer, V. A., & Arkes, H. R. (2009). Preference reversals in evaluations of cash versus non-cash incentives. *Journal of Economic Psychology, 30*(6), 859–872.

Shaw, J. D., & Gupta, N. (2015). Let the evidence speak again! Financial incentives are more effective than we thought. *Human Resource Management Journal, 25*(3), 281–293.

Shugan, S. M. (1980). The cost of thinking. *Journal of Consumer Research, 7*(2), 99–111.

Shurchkov, O. (2012). Under pressure: Gender differences in output quality and quantity under competition and time constraints. *Journal of the European Economic Association, 10*(5), 1189–1213.

Sittenthaler, H. M., & Mohnen, A. (2018). *Cash, non-cash or mix? Gender matters!? The impact of monetary, non-monetary and mixed incentives on performance.* Discussion Paper Series in Economics and Management No. 18-17. Retrieved June 11, 2019, from http://www.geaba.de/wp-content/uploads/2018/08/DP18-17-C1-2.pdf.

Smith, J., Restle, H., & Stanger, M. (2015). *15 companies with incredible employee perks.* Retrieved March 05, 2019, from https://www.businessinsider.com/companies-with-amazing-employee-benefits-2015-8?IR=T.

Smith, J. C. (2015). Pay growth, fairness, and job satisfaction: Implications for nominal and real wage rigidity. *The Scandinavian Journal of Economics, 117*(3), 852–877.

Smith, V. L. (1994). Economics in the laboratory. *Journal of Economic Perspectives, 8*(1), 113–131.

Solnick, S. J., & Hemenway, D. (1996). The deadweight loss of Christmas: Comment. *The American Economic Review, 86*(5), 1299–1305.

Spector, P. E., & Fox, S. (2002). An emotion-centered model of voluntary work behavior: Some parallels between counterproductive work behavior and organizational citizenship behavior. *Human Resource Management Review, 12*(2), 269–292.

Spremann, K. (1987). Agent and principal. In G. Bamberg & K. Spremann (Eds.), *Agency theory, information, and incentives* (pp. 3–37). Heidelberg: Springer-Verlag Berlin.

Stajkovic, A. D., & Luthans, F. (2001). Differential effects of incentive motivators on work performance. *Academy of Management Journal, 4*(3), 580–590.

Stajkovic, A. D., & Luthans, F. (2003). Behavorial management and task performance in organizations: Conceptual background, meta-analysis, and test of alternative models. *Personnel Psychology, 56*(1), 155–194.

Staufenbiel Institut (2017). *Die besten Benefits zum Gehalt.* Retrieved December 28, 2018, from https://www.staufenbiel.de/magazin/gehalt/die-besten-benefits-zum-gehalt.html.

Staufenbiel Institut & Kienbaum (2017). *Recruiting Trends 2017: Was HR Verantwortliche wissen müssen.* Retrieved December 28, 2018, from https://www.staufenbiel.de/fileadmin/fm-dam/PDF/Studien/RecruitingTrends_2017.pdf.

Tangney, J. P., Baumeister, R. F., & Luzio Boone, A. (2008). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality, 72*(2), 271–324.

Thaler, R. H. (1999). Mental accounting matters. *Journal of Behavioral Decision Making, 12*(3), 183–206.

Titmuss, R. (1970). *The gift relationship.* London: Allen and Unwin.

Trachtmann, R. (1999). The money taboo: Its effects in everyday life and in the practice of psychotherapy. *Clinical Social Work Journal, 27*(3), 275–288.

van Dyke, M. (2016). *IRF 2016 trends in incentive travel, rewards, and recognition.* Retrieved July 17, 2018, from http://theirf.org/research/irf-2016-trends-in-incentive-travel-rewards-and-recognition/1692/.

Vohs, K. D., Mead, N. L., & Goode, M. R. (2008). Merely activating the concept of money changes personal and interpersonal behavior. *Current Directions in Psychological Science, 17*(3), 208–212.

Waldfogel, J. (1993). The deadweight loss of Christmas. *American Economic Review, 83*(5), 1328–1336.

Williams, M. L., McDaniel, M. A., & Nguyen, N. T. (2006). A meta-analysis of the antecedents and consequences of pay level satisfaction. *Journal of Applied Psychology, 91*(2), 392–413.

WirtschaftsWoche (2017, February 26). Deutsche Bank Vorstand verteidigt Bonuskürzungen. *WirtschaftsWoche.* Retrieved August 13, 2018, from https://www.wiwo.de/unternehmen/banken/deutsche-bank-vorstand-verteidigt-bonuskuerzungen/19443714.html.

Wiseman, E. (2010, October 17). The truth about men, women, and food. *The Guardian.* Retrieved June 14, 2019, from https://www.theguardian.com/lifeandstyle/2010/oct/17/gender-eating-men-women.

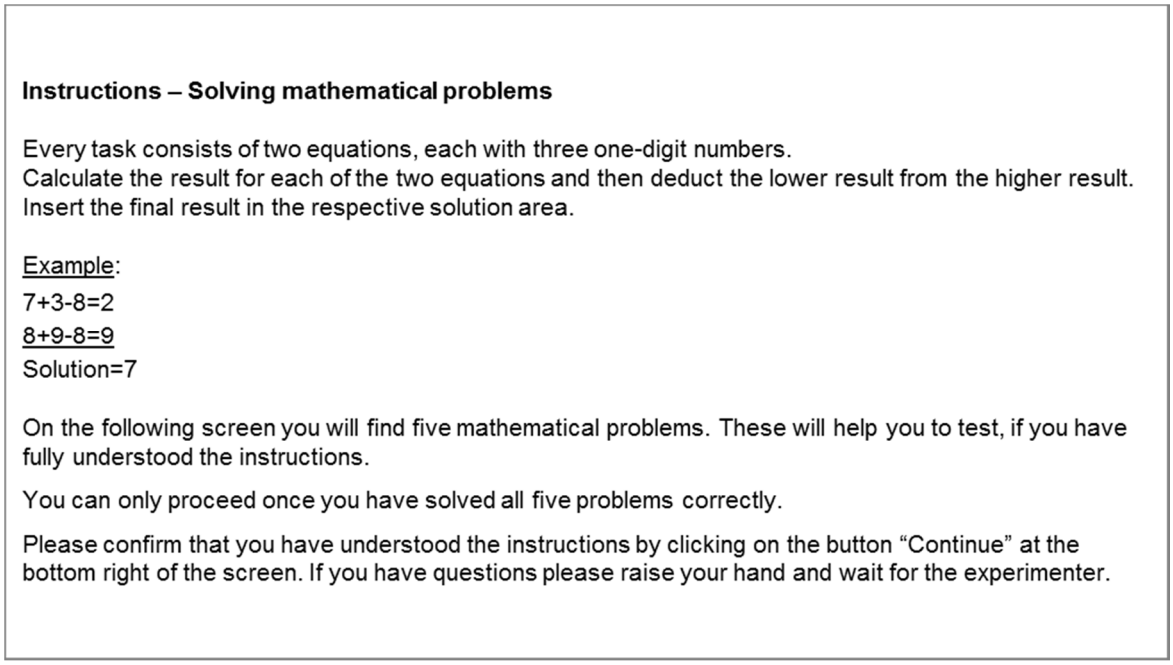
Wiswall, M., & Zafar, B. (2018). Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics, 133*(1), 457–507.

Wooldridge, J. M. (2016). *Introductory econometrics: A modern approach* (6th ed.). Boston: Cengage Learning.

Zepelin, J. von (2017). Was bin ich meiner Firma wert? *Capital, 9*, 28–39.

Appendices

Appendix A: Figures and tables (Chapter 2)



Instructions – Solving mathematical problems

Every task consists of two equations, each with three one-digit numbers.
Calculate the result for each of the two equations and then deduct the lower result from the higher result.
Insert the final result in the respective solution area.

Example:
7+3-8=2
8+9-8=9
Solution=7

On the following screen you will find five mathematical problems. These will help you to test, if you have fully understood the instructions.

You can only proceed once you have solved all five problems correctly.

Please confirm that you have understood the instructions by clicking on the button “Continue” at the bottom right of the screen. If you have questions please raise your hand and wait for the experimenter.

Figure A1: Instructions of mathematical task (translated into English)

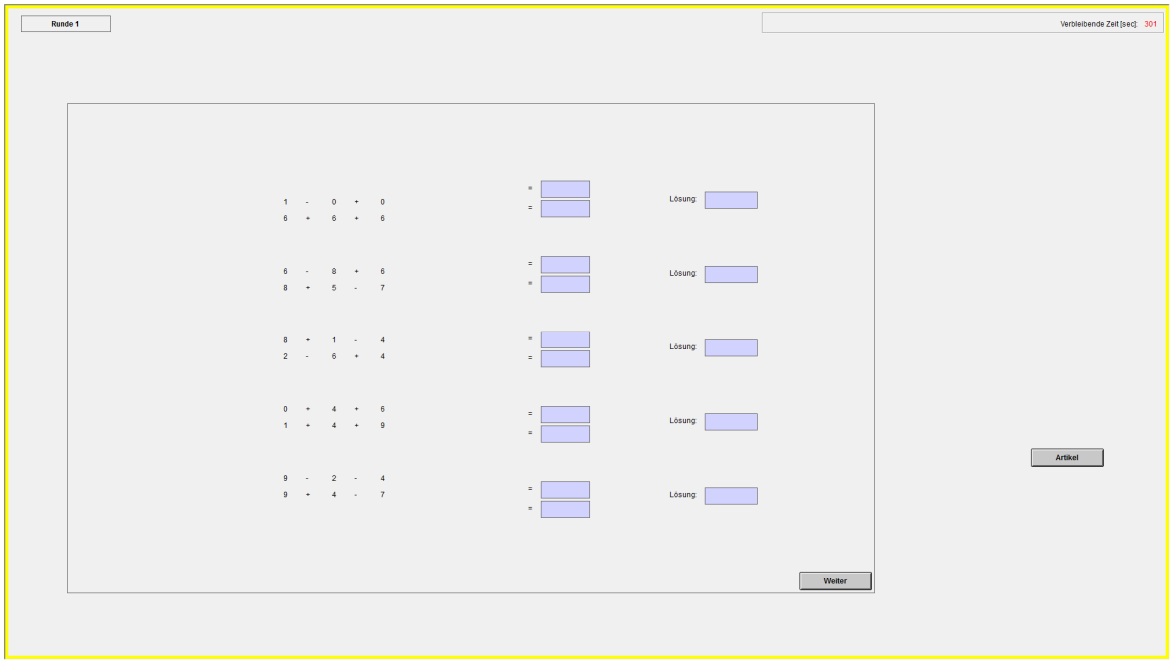


Figure A2: Exemplary screen of working round ("solving mathematical problems")

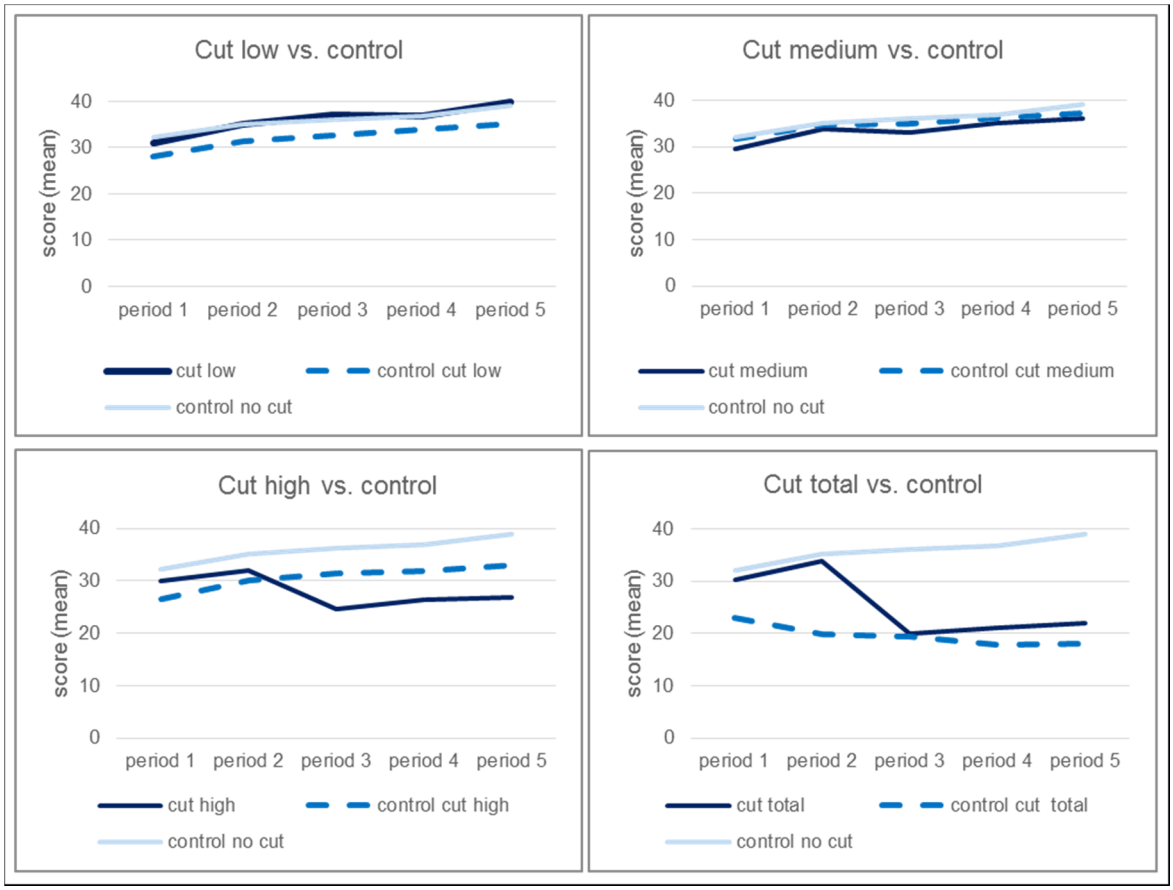


Figure A3: Comparison of performance (*score*) of the cut treatments with respective control treatments

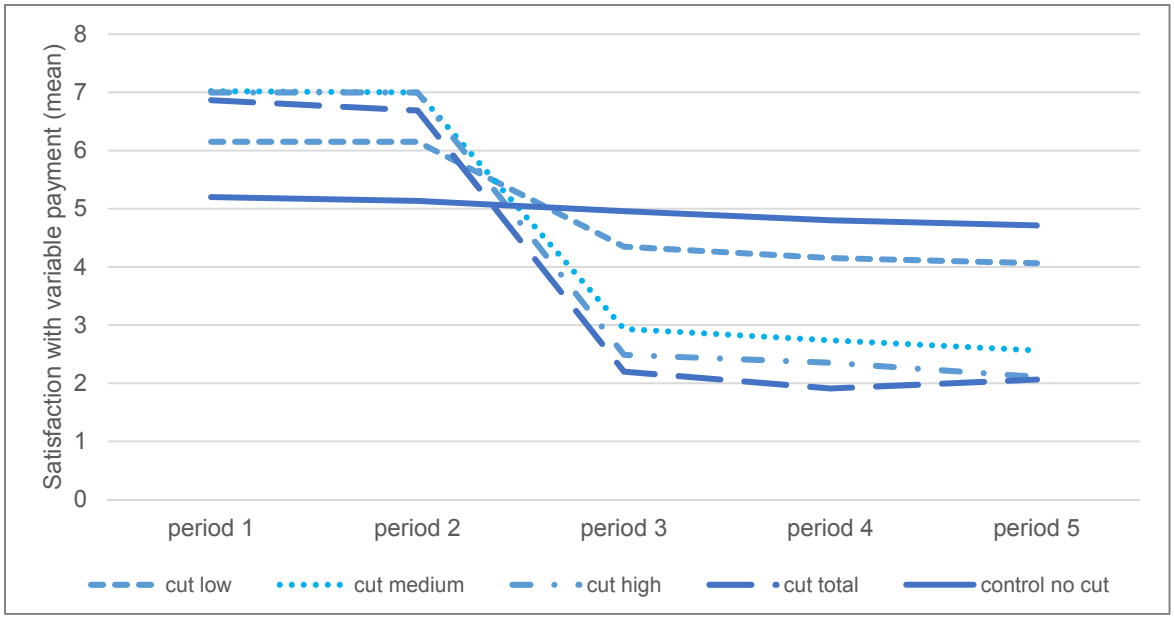


Figure A4: Satisfaction with variable payment over the five working periods and the different treatments

Variable	Description	Cut treatments				Control treatment
		Cut low	Cut medium	Cut high	Cut total	Control no cut
Subjects	Number of subjects	46	46	45	45	45
Gender	Share of women	0.30	0.46	0.38	0.56	0.31
Age	In years	24.33 (2.97)	24.09 (3.40)	23.87 (2.99)	23.31 (2.93)	23.76 (3.18)
Testtime	Mean of time needed to correctly solve mathematical problems in the test period	81.42 (30.11)	101.82 (84.12)	106.67 (57.22)	93.89 (63.28)	100.68 (61.10)
Score1	Mean of number of correctly solved mathematical problems in periods 1,2,3,4, and 5	31.13 (6.86)	29.50 (8.75)	30.07 (7.00)	30.36 (8.27)	32.16 (8.05)
Score2		35.09 (7.28)	33.76 (8.71)	31.98 (7.48)	33.91 (8.06)	35.09 (9.41)
Score3		37.04 (8.07)	32.93 (11.60)	24.69 (15.15)	19.93 (15.70)	36.11 (9.00)
Score4		36.93 (9.24)	35.02 (11.23)	26.36 (15.19)	21.02 (17.57)	36.84 (10.82)
Score5		39.76 (8.08)	35.96 (12.59)	26.73 (16.80)	22.00 (16.81)	39.00 (11.73)
Quality1_p	Ratio of correctly solved problems to solved problems in % in periods 1,2,3,4, and 5	94.73 (4.84)	93.02 (6.36)	94.86 (5.92)	94.42 (5.88)	95.68 (4.22)
Quality2_p		94.35 (5.47)	92.93 (4.78)	94.00 (5.47)	92.52 (6.75)	94.53 (6.05)
Quality3_p		94.07 (5.24)	90.92 (12.66)	89.67 (14.13)	78.01 (29.24)	93.39 (4.70)
Quality4_p		93.83 (4.68)	90.27 (17.19)	87.14 (19.55)	89.41 (17.57)	91.94 (5.45)
Quality5_p		93.96 (4.99)	88.78 (19.19)	88.76 (20.15)	88.81 (14.55)	91.76 (14.62)
Outtime1	Mean of time spent in outside option in periods 1,2,3,4, and 5	7.15 (19.53)	1.96 (4.93)	4.44 (13.64)	2.77 (6.84)	1.49 (4.23)
Outtime2		2.63 (8.66)	2.63 (8.32)	2.72 (7.43)	1.92 (5.26)	2.13 (5.16)
Outtime3		3.82 (13.37)	18.71 (61.96)	71.04 (115.12)	110.10 (128.64)	1.34 (6.89)
Outtime4		6.17 (30.52)	13.27 (47.43)	60.54 (116.14)	107.00 (133.51)	4.47 (20.44)
Outtime5		1.76 (6.95)	13.35 (61.27)	72.82 (119.99)	100.82 (127.56)	8.28 (44.62)
Satisfaction1	Mean of stated satisfaction with variable payment in periods 1,2,3,4, and 5	6.15 (2.85)	7.02 (2.20)	7.00 (2.00)	6.87 (2.68)	5.20 (2.65)
Satisfaction2		6.15 (2.85)	7.00 (2.20)	7.00 (2.01)	6.69 (2.80)	5.13 (2.58)
Satisfaction3		4.35 (2.90)	2.93 (2.52)	2.49 (2.32)	2.20 (2.83)	4.96 (2.50)
Satisfaction4		4.15 (2.87)	2.74 (2.36)	2.36 (2.34)	1.91 (2.66)	4.80 (2.48)
Satisfaction5		4.07 (2.92)	2.57 (2.36)	2.11 (2.38)	2.07 (2.77)	4.71 (2.68)
Fairness1	Mean of perceived fairness of variable payment in periods 1,2,3,4, and 5	5.72 (2.72)	6.30 (2.20)	6.80 (1.95)	6.49 (2.32)	5.11 (2.40)
Fairness2		5.76 (2.73)	6.30 (2.20)	6.71 (1.93)	6.44 (2.32)	5.11 (2.43)
Fairness3		4.41 (2.48)	3.11 (2.17)	2.58 (2.21)	2.64 (2.89)	4.98 (2.38)
Fairness4		4.13 (2.52)	2.96 (2.18)	2.40 (2.13)	2.31 (2.64)	4.80 (2.41)
Fairness5		4.02 (2.60)	2.83 (2.22)	2.20 (2.10)	2.29 (2.63)	4.76 (2.48)

Note: Standard deviations in parentheses.

Table A1: Descriptive statistics

Appendix B: Figures and tables (Chapter 3)

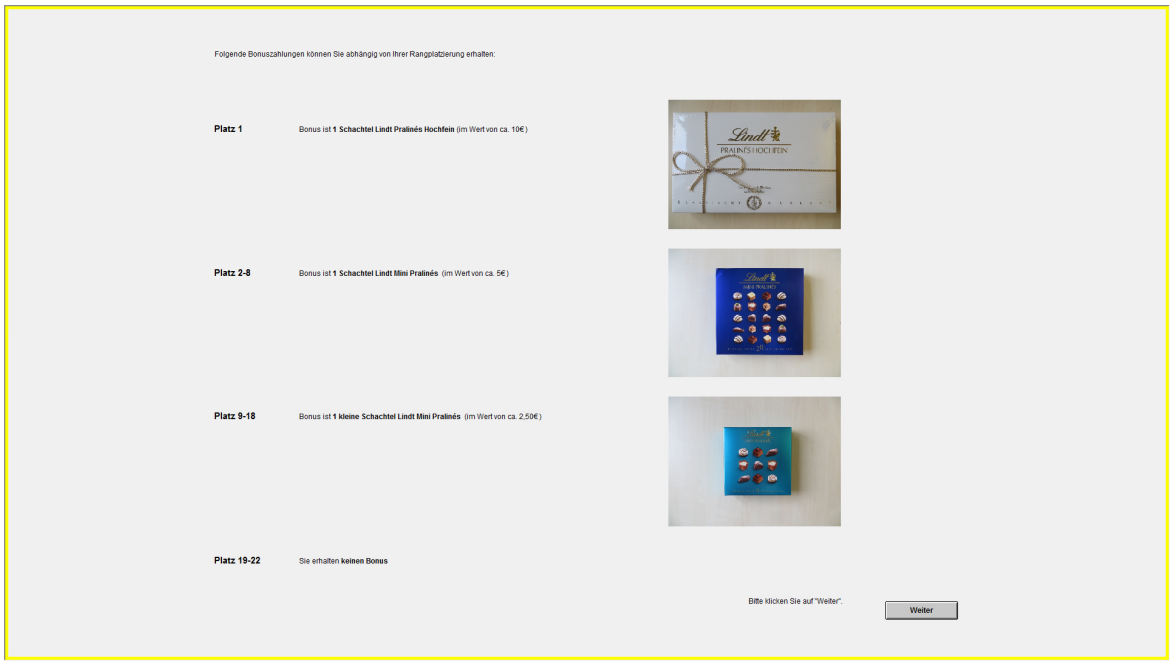


Figure B1: Exemplary screenshot of illustration of prizes (for treatment *nonmonetary*)

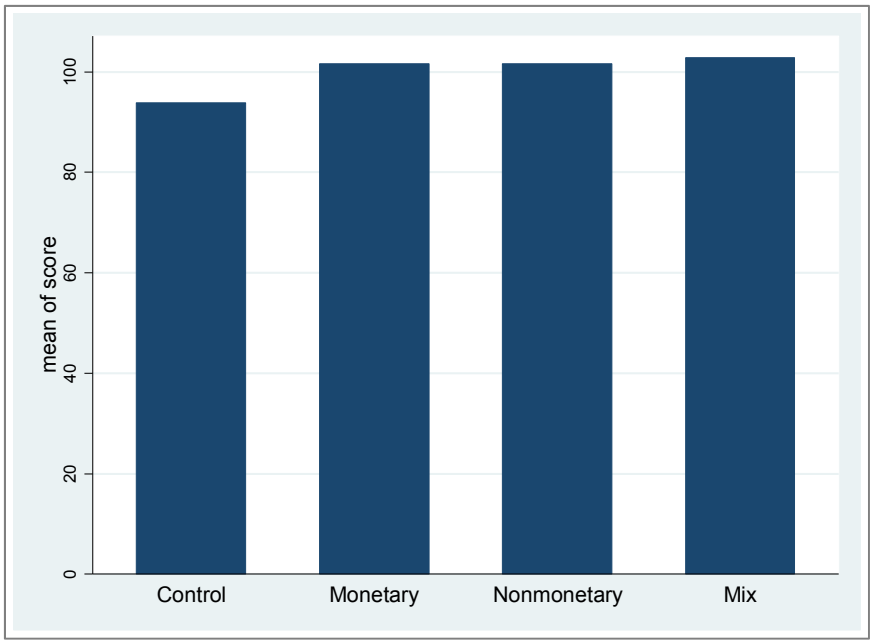


Figure B2: Work performance (mean of *score*) over different treatments

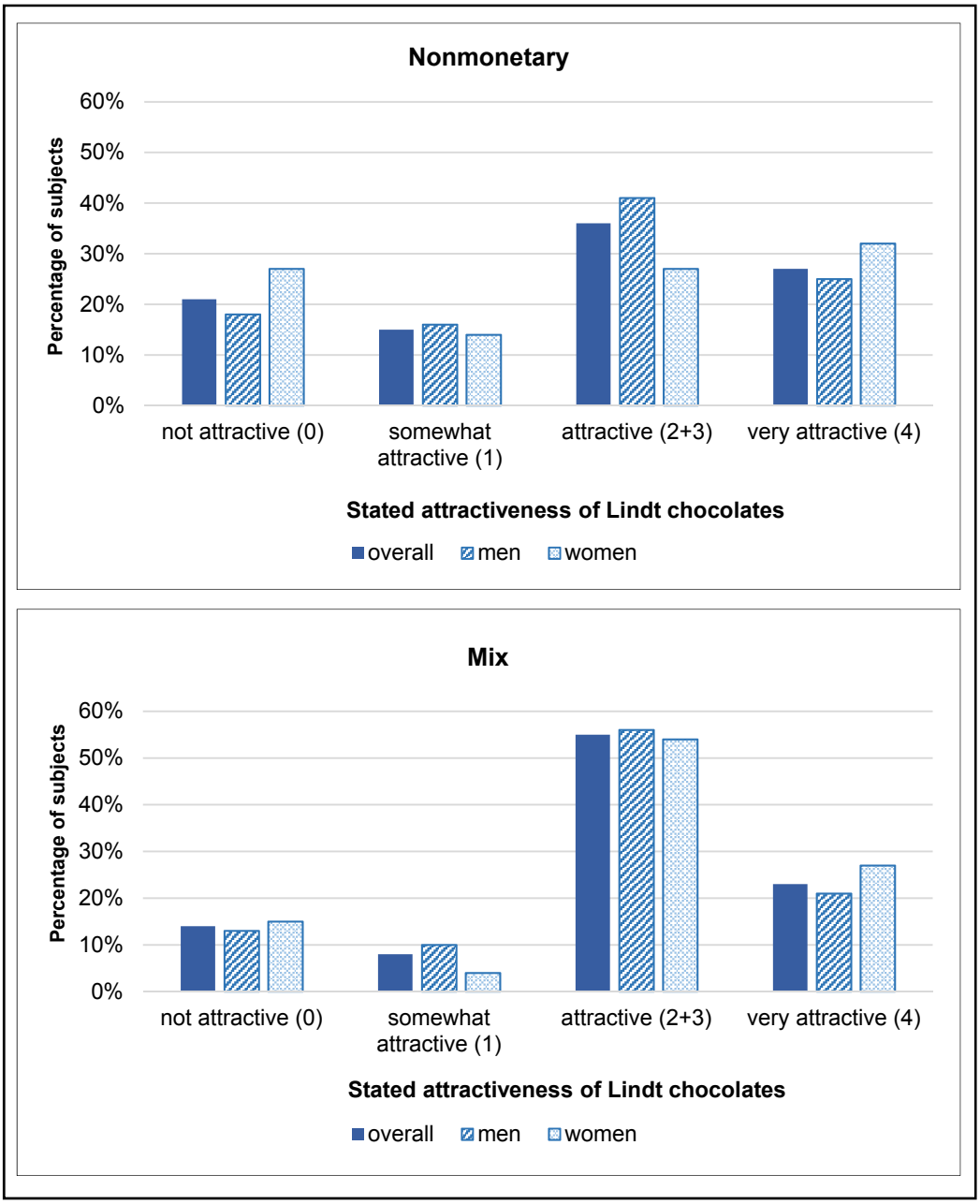


Figure B3: Stated prize attractiveness in treatments *nonmonetary* and *mix*

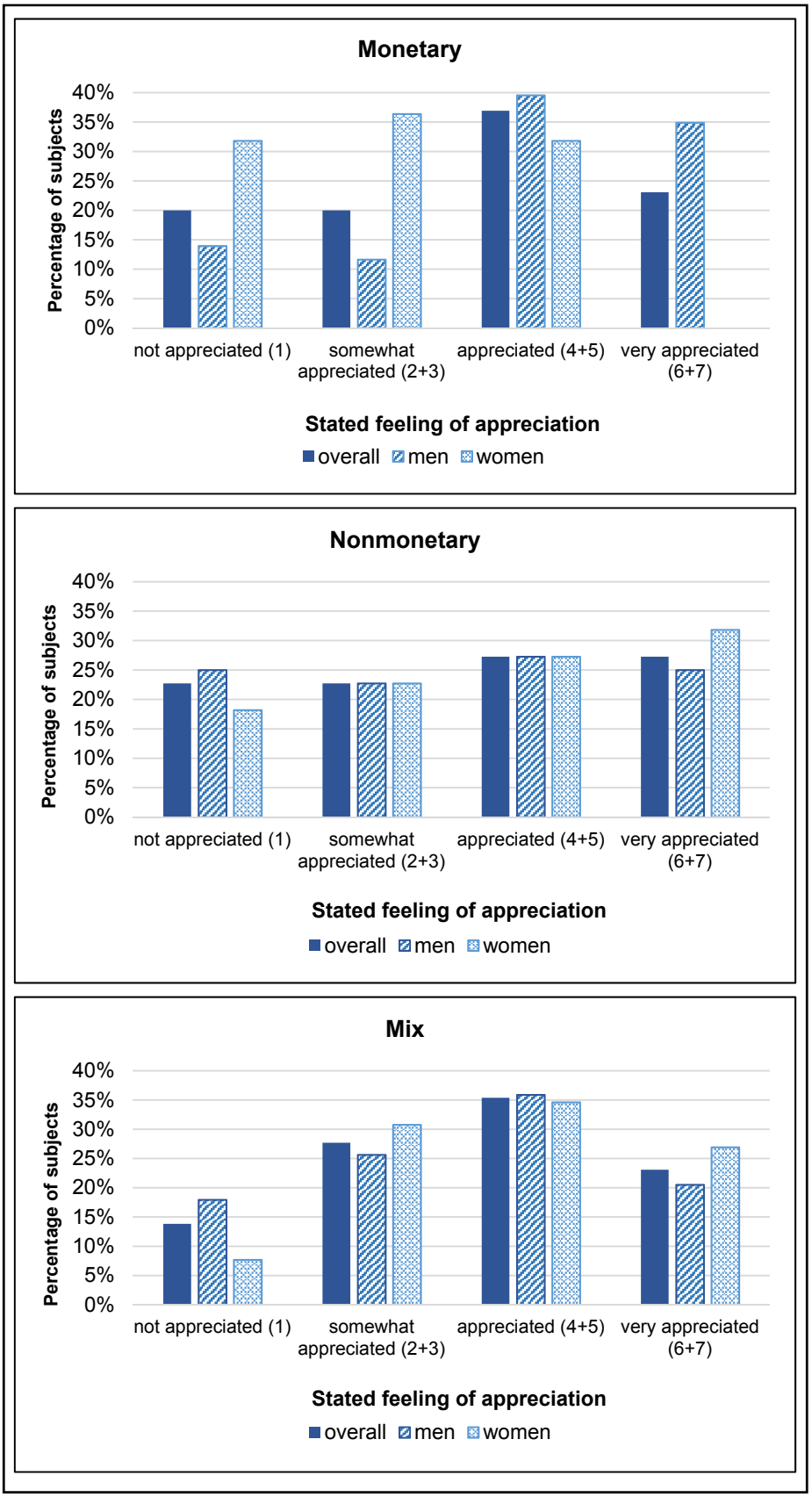


Figure B4: Stated feeling of appreciation in treatments *monetary*, *nonmonetary*, and *mix*

Appendix C: Figures and tables (Chapter 4)

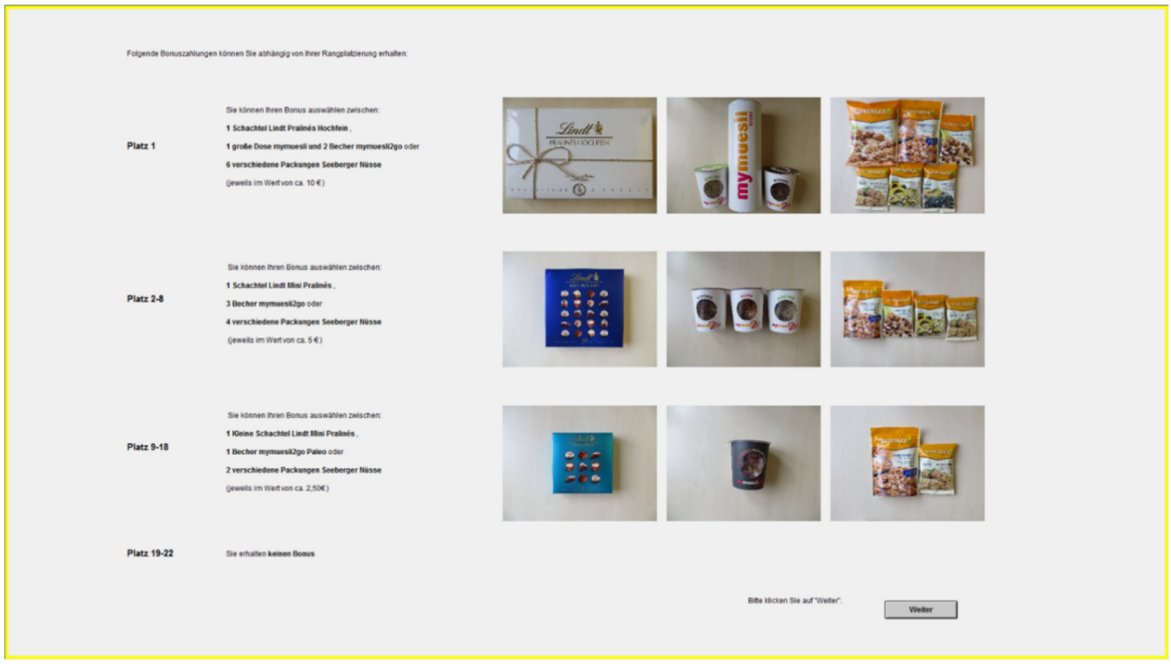


Figure C1: Exemplary screenshot of illustration of prizes in the treatment *cafeteria*

Rational and intuitive decision style scale items (modified version, following Hamilton, Shih, & Mohammed, 2016):

Answers on a seven-point Likert scale (1=strongly disagree to 7=strongly agree):

1. When choosing the bonus, I mainly relied on my gut feelings. (I)
2. I chose the bonus based on intuition. (I)
3. In the selection, I followed my initial hunch. (I)
4. When taking the decision, I weighted the pros/cons of the three bonuses (chocolates, cereals, nuts). (R)
5. I thoroughly evaluated the different bonuses before having made the final choice (R)
6. I weighted a number of different factors, before having decided which bonus I choose. (R)

Figure C2: Rational and intuitive decision making

Treatment	Stated attractiveness of prizes				
	Very unattractive (0)	**Unattractive** (1)	**Neutral** (2)	**Attractive** (3)	**Very attractive** (4)
Chocolate (N=66)	14	10	9	15	18
Cereals (N=65)	15	4	11	16	19
Nuts (N=66)	20	10	12	14	10

Note: Absolute values are presented; prize attractiveness was stated on a five-point Likert scale.

Table C1: Stated attractiveness of prizes in the treatments *chocolate*, *cereals*, and *nuts*

Quality	**Model (1)** Overall	**Model (2)** Overall	**Model (3)** Winner	**Model(4)** Loser
Chocolate (d)	0.00324	0.00438	0.00119	0.00853
	(0.46)	(0.64)	(0.18)	(1.06)
Cereals (d)	0.00459	0.00358	0.000730	0.00642
	(0.67)	(0.53)	(0.11)	(0.83)
Nuts (d)	-0.0165	-0.0161	-0.0173**	-0.0138
	(-1.48)	(-1.43)	(-2.10)	(-0.85)
Gender (=1 if female)		0.001000	0.00331	-0.000931
		(0.18)	(0.55)	(-0.13)
Age (in years)		-0.0000399	-0.00124	0.0000842
		(-0.07)	(-1.65)	(0.13)
Testtime (in sec.)		-0.0000696*	-0.0000789**	0.0000441
		(-1.84)	(-2.50)	(0.90)
Belief (1-4)		-0.0115***	-0.0104**	-0.00944
		(-2.70)	(-2.53)	(-1.58)
Intrinsicmot (1-7)		0.00406**	0.00139	0.00266
		(2.44)	(0.64)	(1.32)
_cons	0.945***	0.959***	1.004***	0.940***
	(177.05)	(51.37)	(42.04)	(39.89)
R^2	0.027	0.084	0.155	0.043
adj. R^2	0.015	0.055	0.094	0.003
N	259	259	120	203

Note: Robust t-statistics in parentheses; significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01; (d) is for binary variable; the dependent variable is the ratio of correctly solved mathematical problems to total problems worked (*quality*); Models (1) and (2) include the 259 subjects of the treatments *control*, *chocolate*, *cereals*, and *nuts*, which worked on the task (number of worked mathematical problems > 0).

Table C2: OLS regressions on work quality (*quality*)

Treatment	Stated feeling of appreciation (mean)		
	Overall	**Winner**	**Loser**
Chocolate	3.74	4.58	3.26
	(2.05)	(2.02)	(1.93)
Cereals	3.62	3.92	3.44
	(1.81)	(1.77)	(1.83)
Nuts	3.47	3.75	3.31
	(1.94)	(2.17)	(1.80)
Cafeteria	3.05	2.83	3.17
	(1.44)	(1.66)	(1.31)

Note: Mean values are reported; standard deviations in parentheses.

Table C3: Stated feeling of appreciation

	Model (1) *Attractiveness of the cafeteria system*	**Model (2)** *motivation*
Justify_chocolate (d)	-1.113**	-0.921**
	(-2.47)	(-2.17)
Justify_cereals (d)	0.367	-0.148
	(1.04)	(-0.25)
Justify_nuts (d)	0.656	0.375
	(1.55)	(0.56)
Gender (=1 if female)	0.658*	0.709*
	(1.80)	(1.84)
Attract_chocolate (1-5)	0.118	0.248
	(0.86)	(1.67)
Attract_cereals (1-5)	0.212*	0.286*
	(1.77)	(1.73)
Attract_nuts (1-5)	0.241*	0.102
	(1.92)	(0.58)
Rational decision taking		-0.325**
		(-2.27)
_cons	3.889***	3.889***
	(6.12)	(5.90)
R^2	0.344	0.353
adj. R^2	0.263	0.260
N	65	65

Note: Robust t-statistics in parentheses; significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01; (d) is for binary variable; the dependent variable in Model (1) is the perceived attractiveness of the cafeteria-style system, in Model (2) the motivation to catalog books; *justify_chocolate*=1 if subject states that it is hard to justify to him- or herself or others choosing *Lindt* chocolates; *justify_cereals*=1 if subject states that it is hard to justify to him- or herself or others choosing *mymuesli* cereals; *justify_nuts*=1 if subject states that it is hard to justify to him- or herself or others choosing *Seeberger* nuts; *attract_chocolate*=stated attractiveness of *Lindt* chocolates measured on a five-point Likert scale (0=not attractive at all to 4=very attractive): *attract_cereals*=stated attractiveness of *mymuesli* cereals measured on a five-point Likert scale (0=not attractive at all to 4=very attractive); *attract_nuts*=stated attractiveness of *Seeberger* nuts measured on a five-point Likert scale (0=not attractive at all to 4=very attractive); *rational decision taking* measures the decision taking behavior of subjects (intuitive vs. rational) and is calculated as the mean of six items (see Figure C2 in Appendix C, each item from 1=not rational at all to 7=completely rational, intuitive items are reversely coded).

Table C4: OLS estimation results regarding the attractiveness of a cafeteria-style system and motivation (scenario survey)

Appendix D: Contribution to working papers

Working paper 1 (Chapter 2)

I was responsible for the development of the research questions and the design of the experiment. Furthermore, I planned and was in charge of the experimental sessions in the laboratory, conducted the data analysis largely independently, and wrote the most part of the article.

_____

Hanna M. Sittenthaler (lead author)

_____

Alwine Mohnen (co-author)

_____

Virginia Herbst (co-author)

Working paper 2 (Chapter 3)

I developed the research questions and the design of the experiment. I planned and was in charge of the experimental sessions in the laboratory, conducted the data analysis largely independently, and wrote the most part of the article.

_____

Hanna M. Sittenthaler (lead author)

_____

Alwine Mohnen (co-author)

Working paper 3 (Chapter 4)

For this experimental study I was responsible for the development of the research questions and the design of the experiment as well as the design of the scenario survey. I planned and was in charge of the experimental sessions in the laboratory. Furthermore, I was responsible for conducting the survey. I analyzed the data largely independently and wrote the most part of the article.

_____

Hanna M. Sittenthaler (lead author)

_____

Alwine Mohnen (co-author)