Labor Market Institutions and Firm Policy

Theoretical and Experimental Insights

Zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften
(Dr. rer. pol.)
von der Fakultät für Wirtschaftswissenschaften des
Karlsruher Institut für Technologie (KIT)

genehmigte

DISSERTATION

von

Dipl.-Kfm. Christian Paul

Karlsruhe, 2009

Referent: Prof. Dr. Siegfried K. Berninghaus
Korreferent: Prof. Dr. Hagen Lindstädt
Tag der mündlichen Prüfung: 18. November 2009

Acknowledgements

First of all, I want to thank Prof. Dr. Siegfried K. Berninghaus not only for supervising my thesis, but also for giving me the opportunity to deepen my understanding of game theory and experimental economics while being a member of his chair at the Institute of Economic Theory and Statistics.

Furthermore, I am grateful to Prof. Dr. Hagen Lindstädt who also supervised and critically discussed my thesis and to Prof. Dr. Wolf Fichtner and Prof. Dr. Kay Mitusch for joining the examination board.

I would also like to thank the several co-authors I had the pleasure to work with during the last years.

In particular, I am indebted to Prof. Dr. Werner Güth who enriched my understanding of economics in numerous fruitful discussions and who allowed me to benefit from the resources as well as the hospitality of the Strategic Interactions Group at the Max Planck Institute of Economics, Jena.

The same thanks go to Dr. Annette Kirstein and Prof. Dr. Roland Kirstein, both working at the University Magdeburg as well as to Dr. Christian Hoppe, and, again, to Prof. Dr. Siegfried K. Berninghaus.

All joint works done with these researchers directly and indirectly contributed to the quality of this thesis. I will detail which parts of this thesis rely on co-authored working papers in the following *Preface*. All remaining errors are my own.

Moreover, I am more than thankful to all my current and former colleagues at the chair of Prof. Dr. Berninghaus. Foremost, I owe a debt of gratitude to Sabrina Bleich, Christian Hoppe, Ralf Löschel, and Marion Ott for all the enriching discussions of my scientific work, for the excellent (working) atmosphere, and, most importantly, for becoming dear friends during the last four years. I also want to thank Prof. Dr. Karl-Martin Ehrhart for his willingness to share his knowledge about statistics, experimental economics and snooker with me.

Finally, I want to try to express my gratefulness to my family, although it is certainly impossible to fully do this in writing. Without the unconditional loving support of my brother Mario Paul and my parents, Wilma and Manfred Paul, this thesis would have never been written.

Contents

Preface **1**

1 Introduction **2**

1.1 Labor market experiments . . . 4
1.2 Framing of experimental instructions . . . 8
1.3 Statistical tests . . . 9

2 Spillover effects of minimum wages **11**

2.1 Introduction . . . 11
2.2 Related literature . . . 14
2.2.1 Minimum wage experiments . . . 14
2.2.2 Relative income studies . . . 17
2.3 The relative income model . . . 20
2.3.1 The minimum wage game . . . 21
2.3.2 Solution design and income comparison utility functions . . . 22
2.3.3 Basic assumptions . . . 24
2.3.4 Equilibria without minimum wage . . . 27
2.3.5 Equilibria with minimum wage . . . 28
2.4 Experimental design . . . 33
2.5 Hypotheses . . . 36
2.6 Experimental results . . . 39
2.6.1 The minimum wage treatment: General results . . . 39
2.6.2 The minimum wage treatment: Wages . . . 41
2.6.3 The minimum wage treatment: Threshold wages . . . 48
2.6.4 The control treatment . . . 52
2.7 Conclusion . . . 57

3 Employment protection and bullying **59**
3.1 Introduction . . . 59
3.2 Model . . . 62
3.3 Experimental design . . . 66
3.4 Hypotheses . . . 68
3.5 Experimental results . . . 71
3.5.1 General results . . . 72
3.5.2 Effort levels . . . 74
3.5.3 Sanctions . . . 79
3.5.4 Payoffs . . . 90
3.6 Conclusion . . . 92

4 Downsizing the labor force **93**
4.1 Introduction . . . 93
4.2 The principal-agent model . . . 95
4.3 Experimental design . . . 98
4.4 Hypotheses . . . 101
4.5 Experimental results . . . 102
4.5.1 Downsizing decisions . . . 103
4.5.2 Treatment AH . . . 104
4.5.3 Treatment AL . . . 109
4.5.4 Treatments AH and AL . . . 111
4.5.5 Treatment(s) UH (and AH) . . . 112
4.6 Conclusion . . . 114

5 Summary and conclusion **116**

Bibliography **118**

A Experimental instructions **133**
A.1 Spillover effects of minimum wages . . . 133
A.2 Employment protection and bullying . . . 136
A.3 Downsizing the labor force . . . 141

B Theoretical addenda – Spillover effects of minimum wages **145**
B.1 Wage ordering 145
B.2 Comparative statics 147
B.3 Total differential, utility of worker L 148

C Statistical addenda – Downsizing the labor force **150**
C.1 List of tests for differences between treatments AH and UH 150
C.2 Logistic regression results, treatments AH and AL 151

List of Figures **155**

List of Tables **156**

List of Abbreviations **158**

Preface

I want to emphasize here that two chapters of this thesis are in large parts identical to current versions of two research papers written with co-authors.

A single-authored analysis of minimum wage spillovers in Chapter 2, is followed by a discussion of the relation between employment protection legislation and the bullying of workers in Chapter 3. The origin of this latter chapter is a working paper I wrote with Annette Kirstein, Senior Lecturer at the University Magdeburg, and Roland Kirstein, Professor at the University Magdeburg. Chapter 4 asks whether less profitable firms fire unproductive workers more often than profitable firms do. This analysis is based on joint work with Werner Güth, Director of the Strategic Interactions Group, Max Planck Institute of Economics, Jena.

I am grateful to these co-authors for permitting me to use our joint work as part of this thesis.

For reasons of consistency, I will use first-person plural personal pronouns in all subsequent chapters – regardless of whether the respective chapter bases upon single-authored or co-authored research papers.

1 Introduction

Labor markets and the questions whether and how politics should regulate them are regularly discussed topics in public discourse across the world. In 2009, for example, changes in minimum wage legislation have been demanded and implemented in the USA, Europe, and Australia, employment protection laws have been an issue in Germany prior to the election of the German Bundestag, and (mass) layoffs caused by the financial crisis have been frequently discussed in many countries (see, e.g., The Australian (2009), Washington Post (2009), or Die Zeit (2009a) for details on minimum wage laws, FAZ (2009), Spiegel (2009), or Die Zeit (2009b) for reports on employment protection, and Guardian (2009), New York Times (2009a), or Times (2009) for articles about layoffs).

With our thesis, we want to enrich the knowledge about these three aspects of labor markets: in Chapters 2 and 3, we theoretically and experimentally analyze the labor market institutions of minimum wages and employment protection, respectively, before we focus on layoffs in Chapter 4.

We start by giving a short overview of the specific facets of minimum wages, employment protection, and layoffs investigated in this thesis. Then, we explain where our work can be positioned in economic literature, before we conclude this chapter by discussing some general questions regarding the experimental and statistical methods we use.

It has been argued by theoretical and empirical field studies that the introduction of minimum wages not only increases the wages of the workers who earned less than this minimum wage before its introduction, but also the wages of workers who were paid more than the minimum wage already. This effect of minimum wages on wages of high-income workers is named *minimum wage spillover effect* or just *minimum wage spillover* in literature. In Chapter 2 we evaluate the current status of research on minimum wage spillovers and propose a new model relying on relative income preferences only. The narrow focus of our model gives us the opportunity to design and implement a laboratory experiment that can distinguish between the various causes for minimum wage spillovers discussed in literature.

In Chapter 3 we deal with another labor market institution, namely employment protection. Here, the relationship between employment protection legislation and the bullying of workers is analyzed. Our starting point is a theoretical and empirical field study by Wasmer (2006) that has suggested that employers might start bullying their workers when employment protection laws disallow them to use firings as an alternative disciplinary incentive. We design a laboratory experiment to test this prediction. In addition, we also briefly investigate *probation period effects* in this chapter, i.e., analyze whether workers spend more work effort in the beginning of their employment where their employers are still allowed to fire them (so-called *probation periods*) than afterwards when firings are not admissible anymore.

We pick up probation period effects and minimum wages in Chapter 4 again, which links this chapter to the preceding ones, but here the focus lies on firms' firing policies. In a laboratory experiment, we investigate the question whether firms that operate highly profitable will lay off their unproductive employees less often than firms that are not very profitable.

Finally, Chapter 5 summarizes the main results of our thesis and proposes starting points for future research.

As indicated above, we use a series of three laboratory experiments to derive insights about labor markets. Therefore, we think this thesis should be seen as a contribution to experimental economics as well as to labor economics.

Of course, experimental economics does also deduce concrete predictions theoretically. However, it varies from study to study whether the emphasis lies on the theoretical deductions or the experimental parts. The same holds for the chapters of this thesis. While in Chapter 2 dealing with minimum wages the theoretical deductions and their experimental implementation are both extensively discussed, slightly more emphasis is put on the experimental parts in Chapters 3 and 4.

We are confident that the series of three theoretical models and laboratory experiments presented here, offers some new insights and, thereby, contributes significantly to the understanding of labor markets. To the best of our knowledge, the concrete questions about labor market institutions we try to answer in Chapters 2 and 3, namely whether minimum wage spillovers exist and whether bullying increases with stricter employment protection, have not been analyzed experimentally before. We are also not aware of another theoretical model focusing solely on relative income preferences to explain minimum wage spillovers. The question we want to answer in Chapter 4 of this thesis is also the main focus of the experimental study by Fischer et al. (2008). However, these authors analyze firms' firing policies in a stylized ultimatum game setting, while we theoretically and experimentally apply a richer principal-agent labor market setting.

Although empirical field studies (henceforth, simply called empirical studies) are still the tool dominantly used in labor economics, labor market experiments have attracted more and more interest in economics during the last years. While, for instance, the standard textbooks on experimental economics like Davis and Holt (1993) or Kagel and Roth (1995) do not feature own chapters dedicated to labor markets, more recent surveys and economic dictionaries do (see, e.g., Gächter and Fehr (2002), Falk and Fehr (2003) or Falk and Gächter (2008)). The trend is also highlighted by the large amount of articles dealing with labor market experiments recently published in highly ranked, renowned journals (see, e.g., Brandts and Charness (2004), Charness (2004), Riedl and Tyran (2005), Brandts and Cooper (2006), Falk et al. (2006), Falk and Huffmann (2007), Fehr and Schmidt (2007), or Healy (2007)). Since we also apply the experimental method to analyze labor markets, we think it is appropriate, if not necessary, to present a general overview of the pros and cons of this approach here already.

1.1 Labor market experiments

The sole purpose of the study by Falk and Fehr (2003) is to deliver a comprehensive overview of the advantages and most common objections against labor market experiments. Consequentially, our summary relies on similar arguments, but regroups and occasionally expands them and adds some newer literature. Similar, yet shorter discussions can be found in Gächter and Fehr (2002) or Falk and Gächter (2008).

We primarily compare the qualities of laboratory experiments to those of empirical studies, since the latter are dominantly used in labor economics. Most often the pros and cons of laboratory experiments, in general, are the same as those of labor market experiments. Thus, we not always make this distinction below.

In our view, every advantage of experiments in comparison to empirical studies has its origin in the (comparatively) strong control over the environment. This not only allows a) to implement institutional changes truly, or at least almost, ceteris paribus and b) to focus on a specific class of effects excluding confounding influences, but also c) to repeat the experiment under (almost) identical settings to strengthen the results' robustness, and, finally, d) to compare results between settings that differ more fundamentally. These latter settings are typically called *treatment conditions* in experimental economics. We now shortly discuss each of these advantages and occasionally use our own minimum wage experiment for exemplification.

Let us start with advantage a). Although, empirical studies can also analyze the effects of the introduction of a minimum wage, they obviously cannot guarantee that nothing else changed during the introduction (like the global economic situation, for example) to the same degree as one can guarantee it in a computer laboratory.

But even if they could, it is still harder for empirical studies to discriminate between causes for results, since confounding influences or statistical noise might play a role. We want to exemplify this advantage b). For minimum wage spillovers, a couple of different effects have been discussed as potential causes in literature, e.g., substitution effects. The basic idea for substitution effects in this context is that if workers of different skills are substitutable from the firms' point of view and the minimum wage increases the relative costs of low-skill workers, then the demand for high-skill workers should increase. This should, consequentially, increase the wages of high-skilled workers, i.e., a minimum wage spillover should follow. The empirical finding that minimum wage spillovers, indeed, occur could thus be used as supporting evidence for substitution effects – but also to support all other causes discussed in literature. However, by prohibiting firms to substitute one worker by another, as we did in our laboratory experiment, we are able to de facto exclude substitution effects as a possible cause for minimum wage spillovers and can focus on other causes.

The pure repetition of experiments unter almost identical conditions as described in advantage c) is done very seldom in experimental economics which is quite unfortunate. Exceptions are prominent games as the ultimatum game or the current punishment literature that, e.g., only changed the cultural background of the subject pool, but not the basic rules of the played game (see Herrmann et al. (2008) who rely on Fehr and Gächter (2002)).

Finally, the advantage of treatment conditions is straightforward: while we are able to compare different settings in our minimum wage experiment (one in which a minimum wage is eventually introduced and one in which it is not) and other treatments are straightforward extensions (for instance, treatments in which a higher or a lower minimum wage is introduced), empirical studies can only analyze the one prevailing setting in the region of the real world they are analyzing. They have to rely on historical data or data from various regions (that likely differ in several other aspects) to be able to make comparisons.

The most immediate objection against laboratory experiments is that their results might only be of limited applicability to real world phenomena. The reasons therefore are said to be found in a) the biased subject pool, b) the just-for-fun character of experiments, because stakes are rather low, and c) the artificiality of the computer laboratory that might question the generalizability of results. Although we cannot ultimately dismiss these arguments, we want to give a couple of reasons and studies that suggest that they are probably not very important.

Students were used in this and most other works as subject pool for the experiments. This is frequently criticized. Falk and Fehr (2003) discuss some experimental studies that test for differences between subject pools. One can summarize the results of these studies by stating that in some cases they suggested that there might be quantitative differences in behavior between several subject pools, but qualitative differences did not occur, i.e., effects were sometimes more or less pronounced for students than for other groups, but strictly opposing behavior patterns were not found. In a recent study, Güth et al. (2007) have also presented evidence that student participants' laboratory behavior is quite similar to that in the field.

Basically the same holds for stake levels. Budget constraints alone prohibit that student participants earn a fortune in laboratory experiments, but nowadays they are paid at least as much as they could earn as a student assistant at the university. In our experiments, e.g., average earnings were about € 10.5 per hour, while the respective wage was € 8. Furthermore, the studies cited in Falk and Fehr (2003) find either that higher stake levels do not alter participant behavior at all or only diminish the variety of responses.

Falk and Fehr (2003) speak of *external validity* when discussing the objection we listed as point c). They point out perspicaciously that, on a meta-level, the predictive power of all scientific deductions – theoretical, experimental and empirical – to other real world situations depends on the question whether the used settings are the same as those in the new situation. In particular, this means that the same objection of external validity can be used against empirical studies, since historical cause and effect might be accurately described by them, but this alone does not guarantee predictive power for new, probably different states of the world. While this is certainly a valid point, it might be criticized that it avoids immediately tackling the question of external validity of experimental data. Instead, it points the finger at other methods' weaknesses.

At first glance, the studies by Karlan (2005), Harrison et al. (2007) and Benz and Meier (2008) seem to use the more appropriate approach to analyze the external validity of experimental data. These authors test the predictive power of experimental results of specific games for similar real world settings by comparing participants' lab behavior with real world behavior (or behavior in similar field experiment as in Harrison et al. (2007)). While the results of Karlan (2005) and Harrison et al. (2007) are mixed, i.e., the behavior of some types of participants and some games have rather high predictive power and others have not, Benz and Meier (2008) find more evidence in favor of external validity. However, we want to stress that using these studies to test for external validity has the characteristics of a circular argument, since it inevitably leads to a similar problem of external validity. Even if assuming that the experimental and field settings they compare were, indeed, identical – what seems problematic enough –, these studies have to compare lab behavior in specific games with behavior in specific field settings. Thus, the question of applicability of these findings to other games and other settings arises again. For example, Karlan (2005) analyzes whether behavior in trust and public-good games can be used to predict real world loan payments of poor Peruvian women. An immediate transfer of his results to the question whether our experimental results can be generalized seems challenging to say the least.

Weighing these pros and cons of labor market experiments, we agree with the discreet conclusion Falk and Fehr (2003) draw: it seems fruitful to use laboratory experiments more often and as a complement to the classical method of empirical field studies, rather than as a substitute. Therefore, we also mention the results of empirical studies analyzing similar questions than we do in Chapters 2 and 3 of this thesis. As Levitt and List (2007) point out, one should probably not use laboratory experiment for exact quantitative predictions, but rather to uncover general behavioral tendencies. We are confident that our experimental insights can be used in this way.

The general discussion just presented allows us to be as brief as possible when using similar arguments in the subsequent chapters and prevents redundancies. For the same reasons, we now shortly explain the framing of our experimental instructions and summarize the basic features of the performed statistical tests.

1.2 Framing of experimental instructions

Since Tversky and Kahneman (1981) the question how to frame experimental instructions, i.e., which words one uses to explain the rules of the experiment, is a topic in experimental economics. As a cover story, Tversky and Kahneman (1981) told participants to imagine that the *"U.S. is preparing for the outbreak of an unusual Asian Disease"* and then asked them to choose between *"two alternative programs to combat the disease"*. In one framing, named *Problem 1*, the two alternative programs were described detailing the probabilities of how many people *"will be saved"*. In the other framing, named *Problem 2*, the survival probabilities for both programs were the same, but now the numbers were presented differently. In particular, the fractions of people who *"will die"* were given. The authors found that participants' choices of programs depended on the framing.

Since we are examining problems specific to labor markets, it seems intuitive to use labor market framing. For labor market experiments, the objections against framed instructions referring to *"employer"*, *"employee"* and so on, are that they might overemphasize participants imported views from outside the laboratory. Note that strictly speaking this is a somewhat different kind of framing effect than that in Tversky and Kahneman (1981) whose cover story was framed in both problems as a medical decision of the U.S. (government). Furthermore, in an experiment similar to ours, Fehr and Schmidt (2007) used labor market framing, but, additionally, performed a control treatment using expressions like *"buyer"* and *"seller"* to check whether framing effects would alter their results. Since they found almost no differences between the two treatments, we decided to cautiously frame all our instructions with standard labor market terminology using *"employer"*, *"employee"* and so on.

However, we are not aware of a study investigating the framing impact of the term *"minimum wage"* that might be more likely to exist. Thus, we chose the conservative way to use *"lower bound"* instead of *"minimum wage"* in the experiment discussed in Chapter 2.

1.3 Statistical tests

If not explicitly mentioned otherwise, we will always use parametric tests for cases with $n \geq 30$ and nonparametric statistics for smaller samples or ordinal data. The latter is the standard method used in experimental economics and diminishes the problem that prior normality tests might lower the overall statistical power. Whenever a significance criterion is necessary, it is $\alpha = .05$. For nonparametric tests, we adjust for ties when necessary.

For explorative data analysis in which a distinction between two- and one-sided tests is possible or in cases in which the respective hypotheses allow two- or one-sided testing, we will use two-sided tests as default. This is insofar conservative as it makes it less likely to falsely reject the null hypothesis. We will use one-sided tests for one-sided alternative hypothesis and then explicitly mention their usage, while we sometimes omit mentioning the two-sidedness for reasons of readability.

There is some controversy in literature whether to use Fisher's Exact tests or χ^2-tests with our without Yates-correction when testing for associations in the two dimensions of categorical data given in a 2x2-contingency table (see, e.g., Haviland (1990) for an overview). In the following we will give p-values of all three types of tests. They never differ qualitatively.

For small samples, we use the Wilcoxon paired sample test (also known as Wilcoxon signed rank test) as the nonparametric equivalent to the standard one sample t test (see Randles (2006)). We also use this test for its original purpose: as a nonparametric dependent sample test.

When comparing the population means of two large samples, we revert to the Welch-Satterthwaite independent two sample t test without prior variance checks instead of using the standard t test. This has been proposed by numerous studies (e.g., Moser et al. (1989), Neuhäuser (2002), Ruxton (2006), or Zimmerman (2004)).

Using the Wilcoxon-Mann-Whitney U-test for detecting differences in central tendency between two small samples is criticized when dealing with differently shaped distributions. In these cases robust rank-order tests were proposed and performed slightly better in simulations with known sample distributions (see, e.g., Fligner and Policello (1981), Siegel and Castellan (1988), Feltovich (2003, 2005), or Ruxton (2006)). Since for data like ours it is not possible to tell whether distributions differ or not with certainty, we looked at the occurring distributions and give robust rank-order tests results in addition to U-test results whenever this visual inspection gave reason to believe that the underlying distributions might differ. For robust rank-order tests, we interpolated the p-values delivered by Feltovich (2005) and used the normal distribution as approximation otherwise. As expected, p-values of both tests are always qualitatively identical.

For reasons of simplification, we abbreviate the *Welch-Satterthwaite test* by *WS test*. We use the term *Mann-Whitney U-test (U-test)* when referring to the two independent sample test and *Wilcoxon paired sample test* or just *Wilcoxon test* only when meaning the one dependent sample test. When performing a robust rank-order test additionally to the U-test we give the results of the former in brackets. Then, we abbreviate the usage of both tests with *U-test (rro test)*. In each chapter, we will mention all these abbreviations one time again.

2 Spillover effects of minimum wages

2.1 Introduction

Recent legislative decisions in Australia, Ireland, and the USA show that minimum wages are still a prevailing topic across the world (see, e.g., The Australian (2009), Irish Times (2009b), or Washington Post (2009)). In this chapter, we theoretically and experimentally examine so-called *minimum wage spillover effects*, instead of focusing on employment effects of minimum wages like the majority of other economic studies did.

What is a *minimum wage spillover effect*? For convenience, let us distinguish between two groups of workers in this chapter: the group of workers whose wages are above the minimum wage before its introduction (henceforth, *high-income workers*), and the group of workers who earn less than the minimum wage before it is introduced (henceforth, *low-income workers*). Literature speaks of a *minimum wage spillover effect* or simply of a *minimum wage spillover* when not only the wages of the low-income workers increase after the introduction of the minimum wage, but also the wages of the high-income workers increase. We follow this convention.

If minimum wage spillovers existed, the consequences for lawmakers and their scientific counsels would be straightforward: in this case they should not only consider the direct effects of minimum wages on wages of low-income workers, but also the indirect effects on wages of high-income workers.[1] Since no other experimental paper known to us concentrates on such minimum wage spillover effects and since the controlled laboratory environment can diminish confounding

[1] The paper of Bauer et al. (2008), e.g., is such an empirical study suitable for policy advice that does not deal with minimum wage spillovers. The authors investigate the employment and fiscal effects of several hypothetical minimum wages for Germany and explicitly exclude potential minimum wage spillovers from their analysis.

effects, we are confident to enrich the knowledge on minimum wages with our study. At the end of this section, we will be able to explain that we also contribute to the literature on relative income with our theoretical model.

Economics textbooks have frequently criticized minimum wages for decades, since simple partial market analyses suggested that minimum wages are destroying jobs (or are irrelevant at best). Although some theoretical studies doubted the universality of this reasoning (see, e.g., Stigler (1946), Drazen (1986), Lang (1987))[2], empirical findings for the USA, Australia, the UK, and Continental Europe have questioned it more enduringly (see Katz and Krueger (1992), Card and Krueger (1994, 1995), Dolado et al. (1996), Stewart (2004)). However, there have been methodological controversies about the two studies that can be seen as the extreme points in answering the question whether minimum wages destroy jobs. While, on the one hand, Card and Krueger (1994) refuted the standard textbook prediction that minimum wages increase unemployment, the study of Leigh (2003, 2004a), on the other hand, largely supported conventional wisdom.[3]

In addition, Katz and Krueger (1992) and Card and Krueger (1995) also provided empirical evidence that the introduction of minimum wages in the US fast-food industry not only raised the wages of low-income workers, but also those of high-income workers. They observe that such minimum wage spillover effects are strongest for workers whose wages were only slightly above the minimum wage before its introduction. The empirical parts of the articles analyzing US data sets by DiNardo et al. (1996), Lee (1999) and Teulings (2000, 2003) imply similar minimum wage spillovers[4], while Dickens and Manning (2004) find only small minimum wage spillover effects for the UK.

Minimum wage spillovers are also suggested by non-scientific publications. For instance, the recent raise in the US minimum wage in July 2009 is said to have an

[2]The monograph of Manning (2003) solely deals with monopsonistic labor markets, the most prominent counterexample against conventional wisdom. Rebitzer and Taylor (1995) combine monopsonies with efficiency wage considerations.

[3]Neumark and Wascher (2000) criticized Card and Krueger (1994) for using telephone surveys and got other results for payroll records. Card and Krueger (2000), in turn, confirmed their initial results when broadening their data set. The findings of Leigh (2003, 2004a) were doubted by Watson (2004) and later defended by Leigh (2004b).

[4]A short overview over these studies can be found in Manning (2003).

impact on wages of high-income workers (New York Times (2009b)).

Several causes for minimum wage spillovers have been discussed in theoretical literature. We want to categorize this literature into four classes: a) models with (partial market) substitution effects, b) relative income models, c) general equilibrium models, and d) search and bargaining models.

The basic idea of models with substitution effects for a firm with heterogeneously qualified workers is straightforward. For simplification, let us assume that only two skill groups exist that also differ in income: high-skilled, high-income workers and low-skilled, low-income workers. If workers of different skills are substitutable, then a minimum wage increases the relative costs of low-skill, low-income labor. Thus firms may want to substitute low-skilled with high-skilled workers. This raises the demand for, and eventually the wages of, high-skilled workers, i.e., substitution effects cause minimum wage spillovers.

Relative income models rely on the idea that the employees' work effort depends on their relative position in the wage hierarchy. Grossman (1983) was the first author focusing on minimum wage spillovers. He explained them in a theoretical model incorporating substitution effects and relative income considerations.[5]

Teulings (2000, 2003) broadens substitution effects to a whole economy in his general equilibrium framework to analyze minimum wage spillovers.

The bottom line of the search models of Flinn (2006, 2008) is, roughly spoken, that by introducing minimum wages, the disagreement outcomes of the Nash bargaining solution increase for all different-skilled workers. This not only increases the wages for low-income, but also for high-income workers, i.e., minimum wage spillovers follow.[6]

We will present a relative income model with heterogeneously qualified workers who also differ in income to analyze minimum wage spillover effects. This means that we will solely focus on one of the several causes for minimum wage spillovers discussed in economic theory. This narrow focus not only allows us to limit our theoretical model to the cause for minimum wage spillovers we perceive best established evaluating the literature, but also to design an according experiment capable

[5]Summers (1988) relies on relative and efficiency wages to analyze unemployment and briefly touches minimum wages.

[6]Some other search models can be found in literature (see, e.g., van den Berg (2003)).

of testing this specific cause. Furthermore, by relying on a relative income model we inevitably also add to the large body of work on relative income. Since Grossman (1983) also uses a relative income model, we will discuss where his approach fundamentally differs from ours when describing our model.

To the best of our knowledge, the works of Brandts and Charness (2004) and Falk, Fehr and Zehnder (2006) are the only experimental studies on minimum wages. Contrary to our study, both studies do not investigate minimum wage spillovers between heterogeneously qualified workers differing in income. Note that Falk, Fehr and Zehnder (2006) confirm the hypothetical value of our research topic by shortly discussing it in their concluding remarks.

Our main research questions are: 1) Does a relative income model predict minimum wage spillovers?, 2) Do minimum wage spillovers occur in an experiment?, and 3) If minimum wage spillovers are found experimentally: are they caused by the relative income considerations discussed in our model?

Our main results are that a) in a theoretic model minimum wage spillovers follow from a rather general set of relative income assumptions, b) minimum wage spillovers also occur in an experiment designed accordingly to our theoretical model, and c) there is evidence that these minimum wage spillovers are mainly caused by the relative income effects discussed in our model.

We proceed as follows: We start by summarizing the experimental studies on minimum wages and the relative income literature most important for our analysis. In Section 2.3, we discuss our four-person minimum wage game and its solution based on a simple relative income model. We give our experimental design of the minimum wage game in Section 2.4 and the hypotheses in Section 2.5. Section 2.6 presents the experimental results, before Section 2.7 concludes.

2.2 Related literature

2.2.1 Minimum wage experiments

We are aware of two other experimental studies analyzing minimum wages: the study by Brandts and Charness (2004) and the study by Falk, Fehr and Zehnder

(2006) (henceforth, FFZ2006).

Insights on minimum wages have not been the sole purpose of Brandts and Charness (2004) who analyzed influences of market conditions in gift exchange games. In an one-sided auction, employers offered wage contracts consisting of a fixed wage to homogeneously skilled employees. Employers could hire one worker at most. This excluded investigating minimum wage spillovers. In a second stage, employees chose effort levels and thereby determined final outcomes. In the treatments with minimum wages, employers were forced to post offers larger than (or equal to) this minimum wage. Amongst other things, Brandts and Charness (2004) found that effort reactions to the same accepted wage were smaller when minimum wages prohibit lower offers. This suggested that workers might have perceived a wage offer only slightly above the minimum wage as rather unfair and thus reacted by spending less effort.[7]

The fairness perceptions caused by minimum wages are the main focus of the study by FFZ2006 whose experiment is more similar to ours. Six firms and 18 homogeneously qualified workers participated in each period of their experiment. Firms' revenues from employing a specific worker did not depend on this worker's effort choice, but were predetermined. All participants knew that each firm could hire up to 3 employees (with decreasing marginal revenues) and that firms were free to offer jobs to 0, 1, 2, or 3 workers, but were limited to unitary wage offers, $w \geq 0$. Due to the unitary wage offers, minimum wage spillovers were impossible by definition. The wages the employers offered were take-it-or-leave-it offers, such that for each single pair of employer and employee the game has the characteristics of an ultimatum game.[8]

The authors used the *strategy method*, i.e., they asked each worker to give a reservation wage (or threshold wage) below which he or she was not willing to work. The workers had to give their reservation wages before they learned the

[7] This is in line with findings in experimental studies on other topics: the set of alternatives for player A seems to be crucial for the perceived kindness and thus the reaction of player B (see, e.g., McCabe et al. (2003) or Falk and Fischbacher (2006) or the more extensive discussion in Section 2.6.2).

[8] In an ultimatum game, a proposer offers a fraction of an amount of money to a responder. If the responder accepts this offer, he and the proposer receive their respective fractions. If the responder rejects, both receive nothing (see, Güth et al. (1982)).

actually offered wages. The reservation wages were insofar a commitment for the workers as they determined their later choices: wage offers below this threshold were automatically rejected. Thereby the *strategy method* gave FFZ2006 the whole strategy profile of each participant, since it determined his or her decision for each possible wage – for hypothetical ones as well as for the one eventually offered. We will employ a similar method in our experiment.

After eliciting the threshold wages via the *strategy method*, FFZ2006 distinguished workers into three groups: a group with low, a group with medium and a group with high reservation wages. In each period, the random matching guaranteed that firms were matched with one worker from each group. The employers then could offer contracts to these workers or not. This was common knowledge. FFZ2006 used this procedure to fasten adjustment processes.

In the first 15 periods of *sequence I* the game was played as described above (unrestricted phase (*u phase*)), before a minimum wage, m – restricting offers to $w \geq m$ – was introduced and another 15 periods followed (minimum wage phase (*mw phase*)). The authors checked for sequence effects. In particular, *sequence II* reversed the order of phases with and without minimum wage, i.e., it started with a minimum wage and then removed it. This means that each sequence consisted of the same two phases, but in different order.

For *sequence I* the authors found that paid wages in all periods were larger than the game-theoretical prediction of $w = 0$. Wages in the *u phase* were smaller than the minimum wage and smaller than those in the *mw phase*. Moreover, the introduction of the minimum wage led to an increase in paid wages above this new boundary m in most cases. The paid wages were about 8 percent higher than the minimum wage. This was in concordance to reservation wages that in many cases also exceeded 0 before and m after the introduction of the minimum wage.[9]

While paid wages in the phases with minimum wage were quite similar in *sequences I* and *II*, the sequencing largely changes the picture for the *u phases*. In *sequence II*, reservation wages and paid wages after the removal of the minimum wage were significantly higher than in *sequence I* that started with the *u phase*. They were now closer to the level of the removed minimum wage.

[9] Note that FFZ2006 use the term *spillover effect* to describe their finding that wages are increased above the minimum wage, while we use it alternatively.

Overall, these results suggest that minimum wages affect reservation wages and paid wages by a kind of *fairness perception effect* that increases wages a little above the minimum wage.

2.2.2 Relative income studies

Essentially, studies on *relative income* are studies on *relative utility*. The common basis of relative income studies is the assumption that an individual's utility not only depends on the absolute income of this individual, but also on this absolute income in comparison to other individuals' absolute incomes. In literature the terms *relative income (utility function)* and *income comparison (utility function)* are thus often used synonymously.

This relativity is also assumed to exist for many other goods, but income is usually used as a leading example. In the broadest sense, i.e., with relative utility for any kind of good, studies are numerous in many sciences. In the following paragraphs we cannot give a complete overview over all relative income studies, but try to present a rather brief, in parts chronological summary of what we perceive to be the main trends. We primarily focus on economic studies on income (see Diener et al. (1999), Frank and Sunstein (2001), Falk and Knell (2004), Clark, Frijters and Shields (2008) or Senik (2009) for more copious summaries of studies in economics, psychology and sociology).

Early empiricists interested in relative utility at least implicitly equated people's answers to survey questions about life satisfaction (or happiness) with people's utility. Recently, this assumption is criticized, but there are some hints indicating that the connection between happiness scores and true utility is not totally random.[10] The most prominent early works on relative income are those of Duesenberry (1949) and Easterlin (1974). The former intended a reformulation of the theory of saving, the latter observed two seemingly contradictory facts nowadays known as *Easterlin Paradox*: 1) within-country empirical findings suggested that absolute income was a good predictor of a person's happiness (and thus utility), and 2) comparisons

[10] An extensive discussion would go beyond the scope of this study. The equation of happiness and utility is commonly dated back to Jeremy Bentham's definition of utility. Critical discussions can be found in Di Tella and MacCulloch (2006), Kahneman and Krueger (2006), Kahneman and Thaler (2006), Kimball and Willis (2006) or Weinzierl (2006).

between countries or within a specific country at different times refuted this clear pattern: in particular, happiness remained relatively stable in several countries during the middle of the 20th century, although absolute income largely increased. Easterlin (1974, pg. 118) concluded that relative income considerations might solve this puzzle and claimed that *"people tend to compare their actual situation with a reference standard or norm, derived from their prior and ongoing social experience"*. The comparison norm that stems from own prior experience is named *adaptation* in literature (see Clark, Diener, Georgellis and Lucas (2008) for an introduction). We will not further discuss this concept here, but we will briefly pick it up below when interpreting our experimental results.

The main working hypothesis derived from Easterlin's original findings was that a person's utility decreases (increases) for increasing (decreasing) incomes of other individuals. This relationship between own and others' incomes is called *relative income effect*. With this effect, it is easy to construct a situation in which absolute income gains might not (proportionally) increase an individual's utility, as long as the relative income position remains unchanged or even deteriorates. Many theoretical works have assumed that the relative income effect exists.[11]

Since the mid of the 90s a lot of economic studies using happiness surveys as a proxy for utility again were published that empirically tested the relative income effect. One can summarize that the vast majority of these studies confirmed the relative income effect (see, e.g., Clark and Oswald (1996), Neumark and Postlewaite (1998), McBride (2001), Blanchflower and Oswald (2004), Ferrer-i-Carbonell (2005), Luttmer (2005), Weinzierl (2006) or Clark, Frijters and Shields (2008)).[12]

In recent years things have become a little more complicated: While results stayed the same for "Old Europe" and for economies as a whole, a number of studies suggested that for transition economies, the USA and within a specific firm an opposite effect might occur, i.e., own happiness/utility might increase when wages

[11] Frank (1984a,b, 1985) is often cited for revitalizing the idea in economic theory, although Hammermesh (1975), Pollak (1976), and Boskin and Sheshinski (1978) earlier relied on similar ideas. Later examples are works on unemployment (Summers (1988) and Akerlof and Yellen (1990)) or on evolutionary games and agency theory (Rayo and Becker (2007)).

[12] Some empirical works by psychologists doubted the relative income effect or at least its magnitude (see, e.g., Veenhoven (1991) and Diener et al. (1993, 1999)). Medical studies suggested that workers earning (much) less than others are unhealthier due to mental distress induced by the lower relative wage rank (Marmot and Bobak (2000), Deaton (2003)).

of co-workers increase (see, e.g., Galizzi and Lang (1998), Alesina et al. (2004), Senik (2004, 2008), or Clark et al. (2009)). This effect is named *future earnings effect* in literature. The intuition is that, especially in the cases described above, increasing wages among peers might be interpreted as a signal or promise that own earnings are likely to increase in the future, rather than as a threat to one's own relative income position.[13] When discussing our model we will explain why this effect should be only of limited importance in our context.

The number of experimental studies (and related theoretical attempts) on relative income depends on how broad one's view on this topic is. Of course, all experimental work on other-regarding preferences, in particular on inequality aversion, is inevitably connected to relative income.[14] The same is true for the theory of relative deprivation originating in sociology (see Clark and Oswald (1998) for a short discussion). When narrowing the view, the quasi-experimental studies on relative income of Solnick and Hemenway (1998), Johansson-Stenman et al. (2002) and Alpizar et al. (2005) are comparable. These studies used surveys in which participants had to decide between two hypothetical societies: one in which their (or their grandchildren's) absolute income was larger, but their relative position was worse, and one with opposite characteristics. They all found evidence in support of the relative income effect.[15] While Charness and Grosskopf (2001) found rather little concern for relative income when maximizing social welfare was another option for experiment participants, Zizzo et al. (2001) and Brown et al. (2008) reported that their experimental subjects were, indeed, interested in (ordinal ranks of) relative income.

The results of a neuroeconomical study by Fliessbach et al. (2007) also support

[13] Hirschman and Rothschild (1973) earlier discussed comparable ideas on a much more general level dealing with economically developing societies. They argued that future earnings effects might only outweigh relative income effects in the beginning of an economic development when others' gains still are perceived as own welfare promises, but eventually disappear when those promises fail to fulfill.

[14] Fehr and Schmidt (2006) summarize theory and experiments in economics. Walster et al. (1978) present psychological experiments based on the seminal *equity theory* paper by Adams (1965).

[15] Solnick and Hemenway (1998) and Alpizar et al. (2005) additionally distinguished between different goods and found relative position to be more important for some goods (e.g. cars) than for others (e.g. vacation time) with income somewhere in between.

the relative income effect. The authors found that during the (laboratory) experiment they performed, the activity in participants' brain regions associated with positive feelings such as rewards increased with higher relative reward payments.

Falk and Knell (2004) and Senik (2009) investigate theoretically and empirically which groups of other individuals are chosen for income comparisons, a question analyzed in many fields of research since the seminal paper on social comparisons by Festinger (1954). Roughly spoken, the results imply that people tend to compare themselves more to similar others than to more divergent ones.[16]

Let us briefly summarize the most important points of this overview of relative income studies: 1) there is reason to believe that peoples' utility from income to some extent depends on relative income, 2) peoples' utility decreases if others' incomes increase and these increases are not interpreted as own future prospects and 3) such considerations might be of greater importance for people more comparable to each other. We will refer to these findings in the following sections.

2.3 The relative income model

In this section, we first introduce the basic rules of our minimum wage game (Section 2.3.1). It is the blueprint for our experimental design and is constructed such that models relying on substitution effects, general equilibrium models and search and bargaining models are not applicable to derive its solution. We outline our alternative solution design in Section 2.3.2. There, we also introduce the relative income utility functions used in literature. In Section 2.3.3, we discuss our set of rather general assumptions, before we derive the equilibria before and after the introduction of a minimum wage, and compare them in a comparative statics analysis in Sections 2.3.4 and 2.3.5.

From now, we will use the terms *employer*, *firm*, and *principal*, on the one hand, and *employee*, *worker*, and *agent*, on the other hand, synonymously in all following chapters and sections of this thesis for reasons of variation. We will also use the feminine (masculine) form for the principal (agents).

[16] Psychology studies of *social comparisons* often analyze comparisons of abilities, rather than incomes. A nice overview can be found in Wood (1996).

2.3.1 The minimum wage game

In our minimum wage game, workers differ in their production skills, i.e., their skill levels determine productivity. In such a setup, it is intuitive to expect that more productive workers earn more than less productive ones. An existing wage hierarchy is, obviously, the prerequisite for analyzing wage spillovers. As will be shown below, the results of our theoretical model do not critically depend on the existence of the specific wage hierarchy in which high-productive workers earn more than low-productive workers, but also hold for other wage orders, as long as there exists any wage hierarchy at all.

The **minimum wage game** is basically designed as follows. One principal (P) interacts with three workers (agents) who differ in productivity. We define productivity as the value of the good produced by the worker. Just like FFZ2006, we assume productivity to be exogenously given to keep things simple. An interpretation could be that effort choices, indeed, determine output, but firms are able to force each worker to spend his specific maximum effort (through perfect monitoring) when hiring them. One of the agents possesses high productivity (agent H), one medium productivity (agent M) and one low productivity (agent L). The worth of the good produced by agent i is denoted by R_i. It is assumed to be immediately sold. We demand the revenues R_i to fulfill $R_L < R_M < R_H$.

The game is played for $l1 + l2$ periods. In the first $l1$ periods **without a minimum wage** the principal proposes a wage tuple (w_L, w_M, w_H) (with $0 < w_i \leq R_i$) that applies to all agents, i.e., the principal is allowed to offer different wages w_i, but is restricted to one specific wage tuple per period. There are no other restrictions or capacity barriers; principals can hire all three workers. We assume that each agent knows all three wage offers.[17] We think that at least when interpreting the employment decision at the end of each period of our game as the decision a worker reaches after a sum of a few real life working periods this assumption is not far-fetched: eventually, every worker in a real world firm is most likely able to accurately approximate others' earnings and to compare these with his own wage. The additional contract rules are those of three unrestricted simultaneous ultimatum

[17]With the aim of receiving more information, we used a kind of *contingent strategy method* in the experiment, see Section 2.4. This is irrelevant for our model and its solution.

games between the principal and each of the three responding agents. Specifically, for our game this means that each agent is confronted separately with the take-it-or-leave-it offer implicitly given by the wage tuple (w_L, w_M, w_H): If agent i chooses to accept the offer w_i, he earns w_i, irrespective of whether the other agents accept their offers. The principal then earns $R_i - w_i$ from employing i. In case agent i declines the offer, he and the principal earn nothing in this relationship. Each agent decides before knowing whether the other agents accept or decline, i.e, decisions are made simultaneously. A principal's overall payoff is the sum of the payoffs from all his workers. We try to keep things simple by disallowing firms to compete for workers. There are also no other similar productive workers competing for the same job. Also, the firm composition does not change which will allow us to analyze wage spillovers within firms.[18]

The introduction of the minimum wage, denoted m, is assumed to happen surprisingly before the last $l2$ periods, i.e., these periods are played **with a minimum wage**. This minimum wage is assumed to lie above the lowest of the wages in the last of the first $l1$ periods and below the second lowest.[19] The minimum wage applies to all workers. All of these rules are common knowledge, except for the introduction and the deduction rule of the minimum wage.

2.3.2 Solution design and income comparison utility functions

For the minimum wage game outlined in the preceding section, game-theoretical predictions for players whose utility solely depends on (and increases in) their own payoff are straightforward: Irrespective whether players' utilities and payoffs are mapped by an identity function what is often used to generate benchmark predictions in literature or by more elaborate relationships, principals should offer the smallest possible amount to all workers before the minimum wage applies (i.e., either 1 if $w \in \mathbb{N}$ as in our experimental setting or the smallest feasible other offer), and m afterwards. The workers should always accept. However, numerous exper-

[18]This is a major difference to FFZ2006 where firm composition randomly changes each period.
[19]For the case of equal lowest and second lowest wages the minimum wage is assumed to be equally high than both wages.

imental studies since Güth et al. (1982) refuted this (see, e.g., Kagel and Roth (1995) for a survey).

The simple set-up of the minimum wage game outlined before prohibits several of the other approaches discussed in literature from being useful: substitution effects and general equilibrium models are not helpful, since firms cannot substitute low-skilled workers by employing additional high-skilled workers, and search and bargaining models can be ruled out, since there is no wage bargaining in our minimum wage game. Instead, relying solely on a relative income model to derive theoretical deductions seems a plausible approach, since workers in the minimum wage game can compare their incomes with that of their co-workers and the theoretical and empirical studies discussed above suggested the importance of relative income. In summary, the construction of the minimum wage game allows us to focus on relative income preferences to theoretically answer whether minimum wage spillovers occur, and this limitation on relative income will, later on, enable us to exclude confounding effects when experimentally testing our theoretical predictions.

Since Grossman (1983) also used relative income considerations to analyze minimum wage spillovers, we want to emphasize that our minimum wage game fundamentally differs from his setting in several aspects. First of all, we concentrate on one firm and exclude substitution effects that largely drive his theoretical results. Furthermore, unlike Grossman (1983) we do not use the simplifying assumption that the low-skilled workers possess no relative income preferences and are always willing to work for the minimum wage, i.e., in our minimum wage game, labor supply of the low-skilled workers is not infinitely elastic. Finally, in Grossman's model workers choose their effort level, while in our setting they only decide whether to work (with a fixed effort and revenue level) or not.

Before introducing the basic assumptions of our model, let us shortly discuss the concrete income comparison utility functions used in literature and their parameter estimates given by empirical studies, since these insights will be helpful later on.

There are basically two classes of functions used to model relative income in economic literature. Suppose worker i's utility u_i depends on his own wage, w_i, and the average wage of workers in his reference group, $\overline{w}_{-i}$.[20]

[20]We will discuss the exact definitions of reference groups in Section 2.3.3.

In *ratio comparison utility* (RCU) functions, utility depends on the own income and the ratio of own versus others' earnings. A concrete example is:

$$\text{(RCU)} \quad u_i \;=\; w_i^{1-\alpha} \cdot \left(\frac{w_i}{\overline{w}_{-i}}\right)^{\alpha} \;=\; \frac{w_i}{(\overline{w}_{-i})^{\alpha}} \tag{2.1}$$

In *additive comparison utility* (ACU) functions, utility depends on absolute differences between own and others' incomes. An example is:

$$\text{(ACU)} \quad u_i \;=\; (1-\alpha)w_i + \alpha(w_i - \overline{w}_{-i}) \;=\; w_i - \alpha\overline{w}_{-i}\,. \tag{2.2}$$

Both types of functions capture the same idea: α is a measure of an individual's concern for relative income with $\alpha = 0$ representing the standard neo-classical model, in which only absolute wage matters and utility and wages are mapped by an identity function. For $0 < \alpha < 1$, an individual's utility is increasing if the own wage increases, but decreases for average wages rising.[21]

Comparable RCU and ACU functions are used by, e.g., Johansson-Stenman et al. (2002), Alpizar et al. (2005) or Weinzierl (2006) to estimate α. Parameter estimates range from $\alpha \approx .3$ to $\alpha \approx .7$, which suggests that relative standing is rather important. Clark and Oswald (1998) discuss theoretical differences of both functions. Johansson-Stenman et al. (2002) and Alpizar et al. (2005) find only minor empirical differences when comparing parameter estimates and prediction accuracy of RCU and ACU functions.

2.3.3 Basic assumptions

We try to derive main insights without committing to specific utility functions which allows us to derive more general results. Thus, the set of basic assumptions (A.1)–(A.3) discussed in this section is held as general as possible. However, occassionally it will be helpful to use ACU and RCU functions as an illustrative guide. In the following sections, we will add further tie-breaking standard assumptions (A.4)–(A.7) that guarantee the existence of an interior solution equilibrium.

The principal and each worker i are assumed to be utility maximizers.

[21] Note that ACU functions also resemble the functions used in experimental literature on inequity aversion or fairness (see, e.g., Fehr and Schmidt (1999) or Charness and Rabin (2002)).

What else do we assume for workers' utility functions? First of all, note that for his empirical parameter estimations of RCU and ACU functions Weinzierl (2006) defined the reference group for worker i by workers of the same gender, same birth year interval and similar education and in their experimental instructions Johansson-Stenman et al. (2002) and Alpizar et al. (2005) simply speak of *"your grandchild's income"* and the *"average income in society"*. The empirical works on relative income mentioned in Section 2.2.2 use similar approaches (see Ferrer-i-Carbonell (2005) for an overview). Although these definitions of reference group (wages) are not extremely narrow, we think that the heterogenous qualification of workers and the fact that there are only three workers in our minimum wage game has to been taken into consideration. We thus assume worker i's utility, u_i, to depend on his wage, and on each of his two co-workers' wages instead of on an wage average only. While workers are likely to compare wages with those of other workers who only differ in productivity, the studies mentioned in Section 2.2 suggest that workers probably do not compare their income with that of the principal who plays a fundamentally different role in the firm. Formally, we introduce $u_i(w_i, w_j, w_k)$ with $i, j, k \in \{L, M, H\}$ and $i \neq j \neq k$.[22]

The principal's utility, which we call profit for reasons of distinguishability from now on, shall depend on the wages paid to the three workers ($\Pi(w_i, w_j, w_k)$).

We also introduce reservation utilities r_i and, for technical reasons only, a binary variable z_i that takes the value of 1, if worker i accepts a contract offer and 0 otherwise. The workers utility and principal's profit functions are assumed to be twice continuously differentiable. We additionally demand (for all workers i):

$$\textbf{(A.1)} \quad \exists\, r_i \text{ with } 0 \leq r_i \leq R_i; \quad z_i = 1, \text{ iff } u_i(w_i, w_j, w_k) \geq r_i$$

$$\textbf{(A.2)} \quad \frac{\partial \Pi}{\partial w_i} < 0; \quad \frac{\partial^2 \Pi}{\partial w_i^2} \leq 0$$

$$\textbf{(A.3)} \quad \forall i \neq j: \quad \frac{\partial u_i}{\partial w_i} > 0; \quad \frac{\partial^2 u_i}{\partial w_i^2} \leq 0; \quad \frac{\partial u_i}{\partial w_j} < 0; \quad \left|\frac{\partial u_i}{\partial w_i}\right| > \left|\frac{\partial u_i}{\partial w_j}\right|$$

The *reservation utility assumptions* **(A.1)** demand that each worker has some reservation utility that must be reached for him to accept a contract offer. From

[22]Below, we will also introduce modified versions of the standard RCU and ACU functions that incorporate this broader definition of reference group wages.

now on, we will assume this reservation utility to be exogenous and fixed, i.e, it will never alter. We think that this is reasonable within the limits of a firm. When thinking of concrete ACU or RCU functions, e.g., reservation utility can be thought of as representing the threshold level the firm's payment scheme must imply for worker i to accept a job offer. It is apparent that $r_i > 0$ is a technical necessity for RCU functions, since they are by definition always larger than zero for positive wages.

The *profit assumptions* **(A.2)** are standard and imply that the principal wants to set wages as low as possible and that the profit function is concave (or linear).

The last assumptions, **(A.3)**, deal with *marginal utilities*. The first two of them are the standard assumptions of non-increasing positive marginal utilities of own income. In accordance with the empirical and experimental findings discussed in Section 2.2, the third assumption states that own utility will, c.p., decrease (increase), if others' incomes increase (decrease). This is exactly the definition of the *relative income effect.*

Although, we cannot definitely exclude the *future earnings effect* in our model (and experiment), we perceive its occurrence to be unrealistic or at least its magnitude to be negligible. Since participants are heterogeneously qualified and we give them no reason to expect this to change (see also the experimental design below), there is little reason for them to perceive other participants' wages as an proper indicator of own future earnings. Additionally, Clark et al. (2009) found evidence that the importance of the future earnings effect diminishes for workers soon to be retired. Although the players in our model (and experiment), of course, do not necessarily retire afterwards, they know that the game will end after the $l1 + l2$ periods. This should further decrease a future earnings effect, if occurring at all. Overall, we think that we can assume that co-workers' earnings are *"uninformative about the individual's own future income prospects"* which according to Clark et al. (2009) is the exact prerequisite for $\partial u_i / \partial w_j$ being negative.

Finally, (A.3) also demands that worker i prefers an own wage increase over an equal sized wage decrease for worker j. This assumption not only seems intuitive, but is also fruitful for equilibrium characterizations as will be shown below.[23]

[23] The reader may argue that while ACU functions obviously always fulfill this assumption, RCU functions do not necessarily have to. However, as will be shown in footnote 26 below, this

2.3.4 Equilibria without minimum wage

The principal's maximization problem without minimum wage is:

$$\max_{w_L,\ w_M,\ w_H} \quad \Pi(w_L, w_M, w_H) \tag{P1}$$

$$\text{s.t.} \quad \forall i: \quad u_i(w_L, w_M, w_H) \geq r_i \ .$$

With two further assumptions, the solution of (P1) describes a subgame-perfect equilibrium.

Firstly, we assume that (P1) has an interior solution (assumption **(A.4)**). This means that we assume a) the positivity constraints ($w_i \geq 0$) for the wages to be not binding and thus irrelevant in the following, i.e., we refrain from discussing the unlikely extreme case of border solutions, and b), consequentially, also demand agents' aspirations to be below their respective revenues. Then, it immediately follows that all three constraints of (P1) must be binding in equilibrium. If, e.g., only the constraint for worker M was not binding, the principal wants to lower w_M which would make both other constraints non-binding which disqualifies this case as an equilibrium. Analog logic prohibits all other cases than the case with three binding constraints from describing the equilibrium.

We assume that an equilibrium exists. Obviously, all following considerations would be meaningless if no equilibrium existed. Additionally to existence, uniqueness of the equilibrium can be guaranteed by a variety of settings: the most immediate approach, for example, would be to demand the sufficient conditions that either the principal's profit function is strictly concave and the constraint set is convex or the opposite case with concavity and strict convexity. However, we will stick to our approach to be as general as possible and simply assume the equilibrium without the minimum wage, $w^* := (w_L^*, w_M^*, w_H^*)$, to be the unique maximum of (P1). The assumption of an existing, unique equilibrium is denoted assumption **(A.5)**.

The equilibrium w^* is then defined implicitly by the first order conditions of the corresponding standard *Lagrange function* $L(w_L, w_M, w_H, \lambda_L, \lambda_M, \lambda_H)$ where λ_i represents the multiplier for the utility constraint of worker i. It is subgame-perfect in all $l1$ periods due to its uniqueness.

special case hurting the last assumption of (A.3) does not limit the results to be derived at all.

With our set of general assumptions, we cannot characterize the resulting wage order. Since it will become apparent in the remainder that it does not matter what specific wage order occurs, but only that there are preferences establishing similar wage orders before and after the introduction of a minimum wage, we limit ourselves here to assume the intuitive wage ordering, namely the one mirroring the revenue differences:

$$\textbf{(A.6.1)} \quad w_L^* < w_M^* < w_H^* \ . \tag{2.3}$$

Appendix B.1 discusses under which assumptions (2.3) holds.

2.3.5 Equilibria with minimum wage

Following our earlier description of the minimum wage game, now a minimum wage is introduced that lies somewhere between the lowest and the second lowest wage: $w_L^* < m < w_M^* < w_H^*$.

We, again, assume the existence of an unique equilibrium $(w_L^{**}, w_M^{**}, w_H^{**})$ fulfilling:

$$\textbf{(A.6.2)} \quad w_L^{**} < w_M^{**} < w_H^{**} \ . \tag{2.4}$$

This seems only consequent considering the wage profile we assumed in (A.6.1) before. As long as the minimum wage lies below w_M^* assuming the wage hierarchy to change seems far-fetched. However, we want to stress again that although we will proceed with (2.3) and (2.4) holding, all subsequent results are also valid for any other wage profile, as long as the wage ordering stays the same.

The principal's maximization problem now becomes:

$$\max_{w_L,\ w_M,\ w_H} \quad \Pi(w_L, w_M, w_H) \tag{P2a}$$

$$\begin{aligned} \text{s.t.} \quad \forall i: &\quad u_i(w_L, w_M, w_H) \geq r_i \\ \forall i: &\quad w_i \geq m \ . \end{aligned}$$

Finally, let us *ex ante* assume for the time-being that the low productive worker's utility, u_L, after introducing a minimum wage m fulfilling $m > w_L^*$, increases, which allows us to drop his participation constraint (assumption **(A.7)**). This simplification is not only necessary for deriving any comparative static results at all, but is justified in so far as it will be shown below that worker L's utility certainly increases for all meaningful parameters of the two classes of income comparison utility functions discussed in literature. We are now able to state:

Proposition 1:
With assumptions (A.1) – (A.7) unique equilibria w^* and w^{**} exist. Positive minimum wage spillovers occur, i.e., workers M and H earn higher wages in w^{**} than in w^*.

Proof: The first part of *Proposition 1* is fulfilled by assumption. With, (A.1), (A.2), (A.6.2), and (A.7), we can simplify (P2a) in four ways. Firstly, by noticing that the principal will set $w_L^{**} = m$, since there cannot be a local optimum with $w_L^{**} > m$ due to the uniqueness of w^* and the non-convexity of Π. Secondly, by deducing from (A.6.2) that the minimum wage constraints for workers M and H are non-binding, thirdly, by detecting that in optimum the utility constraints of (P2a) for M and H must be binding (due to analog reasons as in the previous section), and, fourthly, by dropping the constraint for worker L.
Overall, the principal then solves this maximization problem:

$$\max_{w_M,\ w_H} \ \Pi(w_M, w_H; m) \tag{P2b}$$

$$\text{s.t.} \qquad u_M(w_M, w_H; m) = r_M$$
$$u_H(w_M, w_H; m) = r_H \ .$$

Thus, by introducing a minimum wage, m, fulfilling $w_L^* < m < w_M^* < w_H^*$ we inevitably end up in the cases for which it is possible to do comparative statics (the cases with $u_L > r_L$, $u_M = r_M$, and $u_H = r_H$).

In the corresponding *Lagrangian function* of (P2b), $L(w_M,\ w_H,\ \lambda_M,\ \lambda_H; m)$, the multiplier λ_i again captures worker i's utility constraint. The four first-order conditions of the *Lagrangian function* implicitly define a set of four equations $F(w_M,$ $w_H,\ \lambda_1,\ \lambda_2;\ m)$. Amongst other things, this gives us w_M^{**} and w_H^{**} as functions of m, $f(m)$. By using total differentials, Cramer's Rule and some simplifications[24], the two interesting comparative static results follow as:

[24]See Appendix B.2 for details.

$$\frac{d\,w_M^{**}}{d\,m} = \frac{\dfrac{\partial u_H}{\partial w_H}\dfrac{\partial u_M}{\partial m} - \dfrac{\partial u_H}{\partial m}\dfrac{\partial u_M}{\partial w_H}}{\dfrac{\partial u_H}{\partial w_M}\dfrac{\partial u_M}{\partial w_H} - \dfrac{\partial u_H}{\partial w_H}\dfrac{\partial u_M}{\partial w_M}} \tag{2.5}$$

$$\frac{d\,w_H^{**}}{d\,m} = \frac{\dfrac{\partial u_M}{\partial w_M}\dfrac{\partial u_H}{\partial m} - \dfrac{\partial u_M}{\partial m}\dfrac{\partial u_H}{\partial w_M}}{\dfrac{\partial u_H}{\partial w_M}\dfrac{\partial u_M}{\partial w_H} - \dfrac{\partial u_H}{\partial w_H}\dfrac{\partial u_M}{\partial w_M}} \tag{2.6}$$

The numerators of (2.5) and (2.6) are both smaller than zero, since a positive term is subtracted from a negative one. Furthermore, both denominators are exactly the same which implies that either both comparative static derivatives are positive or both are negative. Less technically, this means that either the wages of both high-income workers increase and we have the classical, *"positive"* minimum wage spillover effect or both wages decrease and we have the opposite result of decreasing wages, which one could call a *"negative"* minimum wage spillover.[25] This also means that the *relative income effect* alone does not necessarily lead to positive minimum wage spillovers. However, the last assumption of (A.3) is a sufficient, but not necessary condition for the spillover to be strictly positive. It implies that both parts in the second term on the right side of the denominators are larger than their respective counterparts in the first term on the left side. Since the second term is subtracted from the first and all other signs vanish, this makes the denominator negative and thus the derivatives $d\,w_M^{**}/d\,m$ and $d\,w_H^{**}/d\,m$ positive.[26] This means that the wages of the medium and the high productive worker increase, although they do not have to, since the minimum wage is smaller than w_M^* and w_H^*, respectively, have been. *q.e.d.*

[25] As mentioned in the introduction, the economic literature uses the term 'minimum wage spillover' only for positive effects.

[26] In footnote 23 we mentioned unlikely cases in which RCU functions do not fulfill the last assumption of (A.3) – with functions defined like the following equation (2.8) and $w_L^{**} < w_M^{**} < w_H^{**}$, e.g., this is the case if either $w_L \leq \alpha_M w_M$, $w_L \leq \beta_H w_H$ or $w_M \leq \alpha_M w_H$. Then, the denominators of (2.5) and (2.6) are nevertheless always smaller than zero for $w_H w_M(\alpha_H \beta_M - 1) < 0$ which is obviously true as long as either α_H or β_M is smaller unity.

We *ex ante* demanded u_L to increase. What is still left is to give some justification for this assumption. With the total differential of $u_L(m, w_M^{**}(m), w_H^{**}(m))$ or simply the chain rule, asking for u_L to increase formally means:

$$\frac{du_L}{dm} = \frac{\partial u_L}{\partial m} + \frac{\partial u_L}{\partial w_M^{**}} \cdot \frac{dw_M^{**}}{dm} + \frac{\partial u_L}{\partial w_H^{**}} \cdot \frac{dw_H^{**}}{dm} > 0 \; . \tag{2.7}$$

Let us now use illustrative functions to determine in which cases (2.7) holds. As mentioned earlier it seems appropriate to slightly modify the standard RCU and ACU functions for our minimum wage game, since we have only three, heterogeneously qualified workers. We propose the RCU function

$$u_i = w_i^{1-\alpha_i-\beta_i} \cdot \left(\frac{w_i}{w_j}\right)^{\alpha_i} \cdot \left(\frac{w_i}{w_k}\right)^{\beta_i} \tag{2.8}$$

and the ACU function

$$u_i = (1 - \alpha_i - \beta_i)w_i + \alpha_i(w_i - w_j) + \beta_i(w_i - w_k) \tag{2.9}$$

with $i \in \{L, M, H\}$, $i \neq j \neq k$, $\alpha, \beta \geq 0$ and $\alpha + \beta \leq 1$, where α is the parameter measuring the relativity preference for the more similar worker (for worker M this worker shall be the low productive worker L). Then, it can be easily shown[27] that (2.7) is larger than zero for all parameter sets with $(1 - \alpha_i - \beta_i) > 0$ holding for at least one subject i. This means that the low productive worker's utility will, indeed, increase after introducing a minimum wage if there is one individual who is at least slightly interested in absolute income. The empirical findings of Johansson-Stenman et al. (2002), Alpizar et al. (2005) and Weinzierl (2006) suggested that the parameter measuring a worker's preference for absolute income – denoted $(1 - \alpha)$ there – should lie somewhere between .3 and .7. Thus, it seems reasonable to expect that our parameter measuring similar preferences, $(1 - \alpha_i - \beta_i)$, should be larger than 0 as well. We then immediately derive that u_L should increase after introducing the minimum wage.[28]

[27] See Appendix B.3 for details.

[28] A possible further step could be, for example, to add factors into the brackets of the RCU function (2.8) that do not demand for equal wages, but, e.g., demand wages to mirror revenue differences. However, this does not change results. See Appendix B.3 for details.

Most of the empirical and theoretical results discussed above argued that minimum wage spillovers are larger for worker M than for worker H. For our model, deductions about the magnitude of wage spillovers have to be rather speculative due to our set of general assumptions. Comparing (2.5) and (2.6), we derive that the medium productive worker's wage change is larger than the high productive one's, iff

$$\frac{\left|\dfrac{\partial u_M}{\partial m}\right|}{\dfrac{\partial u_M}{\partial w_M}+\dfrac{\partial u_M}{\partial w_H}} > \frac{\left|\dfrac{\partial u_H}{\partial m}\right|}{\dfrac{\partial u_H}{\partial w_H}+\dfrac{\partial u_H}{\partial w_M}}\,. \tag{2.10}$$

By (A.3), the denominators in (2.10) are both positive. The numerators represent the absolute utility loss for worker M and H, respectively, from a wage increase for the low productive worker. The studies discussed in Section 2.2 suggest the medium productive worker to be more affected, since comparisons to others are likely to be more important the more similar the other worker is to oneself. This means that $|\partial u_M/\partial m| > |\partial u_H/\partial m|$ should hold – at least for a wage profile mirroring the revenue differences which we will assume for the next paragraphs. But even when abstracting from the fact that the simplifying equalization of $\partial u_M/\partial w_M$ and $\partial u_H/\partial w_H$ might not hold for sufficiently large wage differences, one would additionally need to require the net utility loss $|\partial u_M/\partial w_H|$ to be larger than (or equal to) $|\partial u_H/\partial w_M|$ to come to a definitive conclusion. This requirement means that the absolute utility loss worker M experiences, if worker H leaps one unit further away, outweighs (or equals) worker H's loss, if M comes closer. Then, the denominator on the left side of (2.10) would be smaller than the one on the right side. General studies on upward and downward comparisons do not give an unambiguous answer to the question whether $|\partial u_M/\partial w_H|$ is larger than $|\partial u_H/\partial w_M|$ or not, especially when considering that we deal with heterogeneously qualified workers (see, e.g., Major et al. (1991) or Falk and Knell (2004) for (short) summaries), but the empirical study by Ferrer-i-Carbonell (2005) finds some evidence for a trend for upward comparisons in Germany.

We thus limit ourselves here to state that if the net utility gain from equal marginal wage changes for both workers has the same effect on both of them, i.e., if the denominators in 2.10 are equal, then wage spillovers should indeed be larger for worker M than for worker H.

Let us now quickly discuss if our model can incorporate the findings of FFZ2006 who argue that minimum wages might change fairness perceptions and thus increase wages of workers immediately affected by the minimum wage. An extension of our model that incorporates this idea is straightforward: If the minimum wage establishes a kind of psychological barrier that the new wage w_L must significantly lie above, this could be caught by substituting m with $m + \delta_m$ ($\delta_m > 0$), i.e., by demanding the new wage to be a little larger than the minimum. The main results are obviously qualitatively unaltered from this modification as long as $m+\delta_m < w^*_M$ holds. As we will discuss in greater detail during our experimental results, there is little reason to expect a similar *fairness perceptions effect* influencing the wages of workers M and H, since they are likely to earn a lot more than the minimum wage before its introduction. This also means that $m + \delta_m < w^*_M$ is not a very restrictive assumption.

In our setting the minimum wage is newly introduced, i.e., it is introduced in a labor market in which previously no minimum wage existed. Nevertheless, it is apparent that the effects we discuss would also apply for the alternative case in which lawmakers raise an existing minimum wage to a new level as long as the minimum wage change fulfills our assumptions. We thus think our results are also important for such changes in existing minimum wage legislation.

2.4 Experimental design

Our experiment basically followed the model, i.e., groups of four participants interacted during the experiment: one principal (P-participant), and the three workers (participants L, M, and H). The roles were assigned randomly before the start of the experiment and stayed the same throughout. Up to five groups of four formed a session with 20 participants.[29] Anonymity was guaranteed by seating players arbitrarily into two laboratories.

[29]Due to a larger number of no-shows than expected from prior experiments, we had to run a couple of sessions with less than 5 groups. Since each group of four did not interact with other groups, this should not be important.

We set the revenues to $R_L = 100$, $R_M = 200$, and $R_H = 300$. Each group interacted for 10 periods. In each period, the principal had to offer a contract consisting only of a fixed wage to each agent, i.e., to propose a tuple (w_L, w_M, w_H). We restricted the offers to positive integers smaller than or equal to the corresponding revenue. As in the model, workers decided whether to accept their offer or not before they knew the other workers' decisions.

In the minimum wage treatment (**treatment MW**), the first 5 periods (**MW1-5**) were played as follows: First, principals offered a wage tuple (w_L, w_M, w_H), then each agent i was informed about the offers to the two other agents and was asked to name the threshold wage t_i ($\leq R_i$) above or equal to which he was willing to accept the principal's offer.

This means we apply a modified *strategy method* in our experiment. The *standard strategy method* is frequently used in experimental economics. It demands from participants to make their decision for all hypothetically occurring cases of the played game (e.g., for all possible wage offers), before the one case that really applies (e.g., the one concrete wage offer) is revealed to them. This gives researches the whole strategy profiles of players, while the *decision method* in which players only decide for the one factual case that applies does not offer these insights.[30]

Asking workers to name their threshold wage knowing the wage offers to the other worker means that we deviate from the standard strategy method and introduce a kind of *contingent strategy method*. It gives us the worker's decision for each of his (up to 300) possible wage offers, but limits this insight to the wages actually offered to the other workers. We think that this is the logical step in a case like ours in which the standard strategy method is theoretically possible, but not practicable: for worker L, for example, the standard strategy method would require to ask his threshold wage for each wage combination (w_M, w_H) – in our setting 60,000 $(200 \cdot 300)$ of these combinations exist.

We admit that using the contingent strategy method might have accentuated participants' genuine disposition for income comparisons (as well as the standard

[30]The advantage of the decision method is that it represents the more natural way of decision making. However, most studies found no (see, e.g., Brandts and Charness (2000), Seale and Rapoport (2000), Oxoby and McLeish (2004), Solnick (2007)) or only negligible behavioral differences (Casari and Vason (2009)) between strategy and decisions methods, while Brosig et al. (2003) report some differences.

strategy method would have), but perceived this a minor sacrifice compared to the broader insights to participants' decision behavior. Moreover, even using the *decision method*, i.e., giving participants the whole wage profile (w_L, w_M, w_H) before their decisions, should have a similar effect, since wage comparisons were still possible. This strengthens the comparative advantage of the contingent strategy method in our situation.

We are not aware of any other study that applies a similar contingent strategy method. Our approach is thus a novelty.

If an agent's threshold was lower or equal than the offer, he earned w_i and the principal $R_i - w_i$, else both earned 0. A principal's overall payoff was the sum of the earnings in all three simultaneous ultimatum games. After periods 1 to 5 (p1-5) there was a short break in which participants got the instructions for the *"second part"* of the experiment.

In the last 5 periods (**MW6-10**) contract offers were further restricted by minimum wages. Instead of pre-committing to a specific minimum wage, the minimum wage was deduced by the contract offers of period 5. Since we are interested in relative changes in wage hierarchies, we feared that *ex ante* defining a specific minimum wage for all groups would question their comparability, since then the minimum wage could lie above, below or between the existing wage profiles. We chose period 5, because wages were likely to increase during p1-5 (see Section 2.5). Let us denote the lowest wage in a specific period with w_{low} and the second lowest wage with w_{s-low}. We set the minimum wage, m, to the integer nearest to the point defined by $w_{low} + .25(w_{s-low} - w_{low})$, i.e., the minimum wage was set between the lowest and the second-lowest wage paid in p5.[31] We additionally restricted it to be not larger than 99 to guarantee meaningful offers to worker L. Participants did not learn this deduction rule, and were only told the specific minimum wage applying to their group. Although we are not aware of other experimental studies that use a similar method of a group-specific treatment condition, we want to stress again that we perceive it as crucial to guarantee comparability of observations in an experiment like ours. Workers' wage thresholds additionally had to fulfill $t_i \geq m$, i.e., we allowed only meaningful threshold wages which is another difference to FFZ2006.

[31] Or at least equal to both in case of equal wages.

We ran a control treatment (**treatment CTR**) to check whether wage increases in p6-10 would also occur when no minimum wage was introduced. If this was the case, one would have to consider that possible wage spillovers detected in treatment MW were maybe not originating in relative income motives. In treatment CTR periods 1 to 5 (**CTR1-5**) were played exactly as MW1-5, but after the break 5 identical periods followed (**CTR6-10**).

At the beginning of each session, participants were randomly assigned to the roles of employers and employees. We gave participants no reason to believe that their productivity might change after MW1-5. On the one hand, stating something like *"your productivities will not change"* would have diminished possible, but anyhow improbable, *future earnings effects*, but, on the other hand, we feared that with this information participants might have started to ask themselves what else would change. Participants did know however, that there was going to be *"another part of the experiment"*. To the best of our knowledge, no study exists that explicitly analyzes the framing effect of the term *"minimum wage"*. We thus chose the conservative way to describe the minimum wage as *"lower bound"* additional to the standard labor market terminology used in our instructions. Representative instructions are given in Appendix A.1.

In our experimental session, 35 groups altogether played MW, 30 played CTR. Before the experiment started, participants had to answer some control questions (at their computer terminal) that checked their proper understanding of the instructions. All sessions were conducted at the experimental computer laboratory in Karlsruhe, in May and June 2009. Participants were students from the Universität Karlsruhe (TU), mainly in business engineering. Average earnings were € 13.25 for about 65 minutes (about € 12.50 per hour).

2.5 Hypotheses

Our main research question is whether minimum wage-spillovers occur or not. For treatment MW, it makes sense to contrast between three kinds of hypotheses to tackle this question: a) the **benchmark hypothesis** with standard rational, payoff-maximizing subjects, b) the **fairness hypothesis** that loosely interprets findings from ultimate game experiments and other behavioral literature, and c) the **relative income hypothesis** based on our relative income model.

The benchmark predictions based on the subgame-perfect equilibria with payoff-maximizing players and *common knowledge* are straightforward.

Benchmark hypothesis (hypothesis I)*: Employers offer the smallest possible wage to all three workers in p1-5 and the minimum wage in p6-10; employees always accept these contracts. Full employment results. No minimum wage spillovers occur.*

Results of ultimatum games doubt these predictions, although most often ultimate game experiments focused on only one game and not three simultaneous ones. In ultimatum game experiments, proposers frequently offered more equal proportions of the pie and responders regularly denied unequal offers (see Kagel and Roth (1995) or Camerer (2003) for comprehensive summaries). The equal split often was the modal observation. Motives like altruism, or fairness preferences (see, e.g., Rabin (1993)) were used to explain these results.

FFZ2006 use the heuristic of about 30 to 40 percent of the total pie size to predict firms' offers in their experiment which is comparable to ours. Offers below this should usually be rejected. Similar approximate estimations are given in meta-studies like Fehr and Gächter (2000b) or Camerer (2003). The rejected offers suggest to expect less than full employment. When focusing on the findings of aspired pie shares, these results do not indicate that minimum wage spillovers should be expected.

Fairness hypothesis (hypothesis IIa)*: In p1-5, employers offer on average about 30 - 40 percent of the revenue R_i to worker i. In p6-10, employers either offer about the same as before or the minimum wage (depending on which value is larger). Less than full employment results and no minimum wage spillovers occur.*

The findings of Brandts and Charness (2004) and FFZ2006 suggest to modify hypothesis IIa a little. They both emphasize that a minimum wage might be perceived as a new wage threshold the principals have to overbid to induce acceptance of contract offers. FFZ2006 find in their setting that wages lie about 8 percent above the minimum wage.

Fairness hypothesis (hypothesis IIb)*: In p6-10, employers either offer about the same as before or more than the minimum wage (depending on which of these two values is larger).*

Due to the rather large productivity differences, intuitive reasoning lets us expect a wage profile mirroring the productivity differences, i.e., $w_L \leq w_M \leq w_H$ should hold, probably with strict inequality. Positive minimum wage spillovers should occur if our assumptions hold and relative income matters.

In every experiment testing a model with common knowledge, one has to consider that it should take a while for employers to accurately predict their workers' reservation wages (and the underlying utility functions).[32] Thus, denials of offers should occur and wage offers should follow an increasing trial-and-error path.

When discussing the experimental results below, we will argue that focusing attention to the wage offers that actually led to employment is most important. These wage offers are the wages that have to be paid in the end. Therefore, the paid wages describe the contracts actually applying and exclude the extreme (trial balloon) offers that were not accepted. We thus formulate the following hypothesis for these paid wages already, although the differences are almost only semantical here. We will extensively discuss whether the experimental results differ between all wage offers (those that led to employment and those that did not) on the one hand, and paid wages only on the other hand.

With regard to paid wages, we feel tempted to predict them to be a little higher than 30 - 40 percent of revenues, because these are the wages that are accepted. We thus use the upper bound of the wage interval proposed by FFZ2006 as benchmark prediction. As discussed before, our basic model does not directly predict wages for worker L to exceed the minimum wage.

Relative income hypothesis (hypothesis III)*: In p1-5, paid wages amount to about 40 percent of the revenues R_i. In p6-10, worker L is paid the minimum wage. The wages of workers M and H increase compared to their wages in p1-5, i.e., there are positive minimum wage spillovers. Less than full employment results.*

Treatment CTR was performed to provide a check whether wages increase in

[32] This is not only true for our work and fairness models, but also for the standard game-theoretic model underlying the benchmark hypothesis if taking the theory to its logical conclusion.

p6-10 anyhow and whether the wage increases in MW6-10, if occurring, are larger than those in CTR6-10. Since CTR6-10 was played with the same rules as CTR1-5, only the parts of the hypotheses dealing with p1-5 apply to treatment CTR.

2.6 Experimental results

In treatment MW the tie-breaking wage of 99 applies in one group, i.e., we have 34 of 35 groups that are identical in so far, as the minimum wage is set equal to $w_{low} + .25(w_{s-low} - w_{low})$. Consequentially, the group violating this rule is excluded from the following analysis.[33] Furthermore, the averages of all wage offers for all workers increase during MW1-5 and, quite surprisingly, during MW6-10. This is illustrated in the following Figure 2.1 whose vertical axis is limited to a reasonable interval.[34] The figure visually confirms that using the wages of p5 to determine the minimum wage was the right decision.

As expected, the condition $w_L \leq w_M \leq w_H$ holds in all 34 groups of treatment MW for period 5, and the two averages MW1-5 and MW6-10. In MW1-5 and MW6-10 it does so with strict inequality in 33 out of 34 cases; in period 5 of MW this is true in 31 cases. In treatment CTR things are even more unambiguous: In period 5 of MW and in MW6-10 the condition $w_L \leq w_M \leq w_H$ holds with strict inequality in all 30 cases. The same is true for 29 groups in MW1-5, and only one group violates even the weaker equality restriction. In this particular group, the principal offers worker M much more than worker H in the first two periods, but then switches to the expected wage profile with strict inequality after noticing that worker H never accepts. Overall, the wage profiles are as expected which shows that the earlier assumptions were not far-fetched.

2.6.1 The minimum wage treatment: General results

For each group of four (the principal P and the three workers L, M, and H) we calculate two averages: the average of, e.g., wage offers before the introduction of

[33] Apart from the quite high wage offers, results for this group do not differ from the other groups. Data on request.

[34] We will do this in all following figures of this thesis for reasons of presentability.

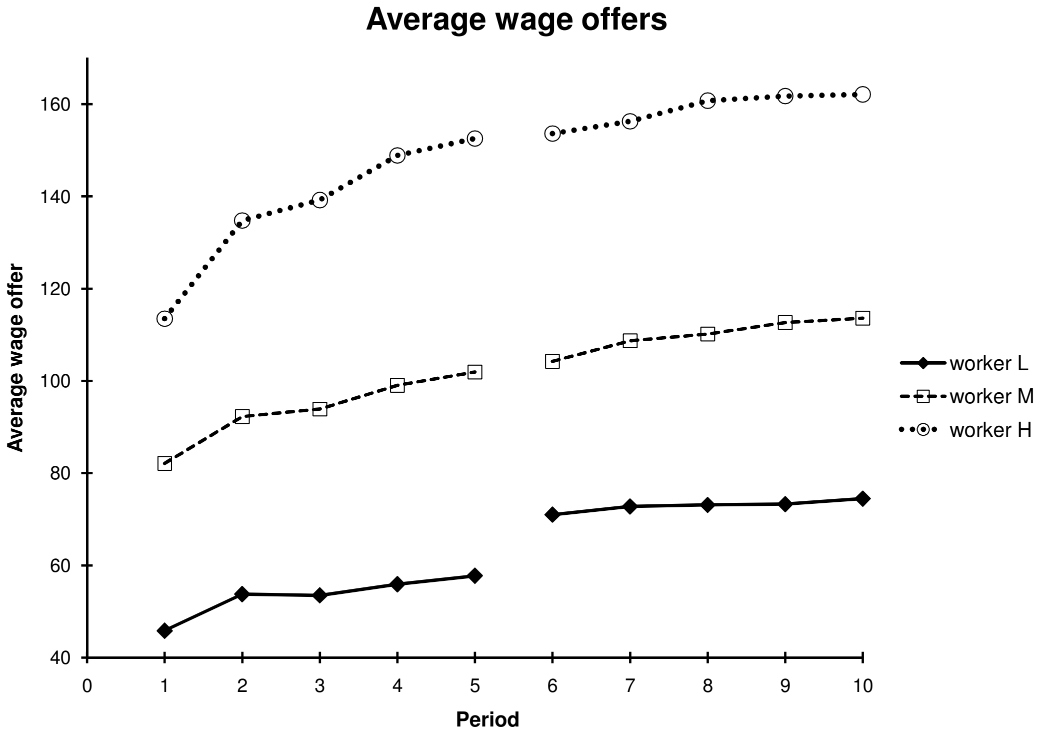

Figure 2.1: Treatment MW, average wage offers

the minimum wage (MW1-5), and the one after its introduction (MW6-10). The averages over all groups in MW1-5 are treated as mutually independent observations, the same holds for MW6-10. The observations after the introduction of the minimum wage (MW6-10) are dependent on the ones before (MW1-5). In Table 2.1 averages of totally hired workers (hrd), the resulting payoffs (Pay_i) and the welfare defined as payoff sum (WF) are given. This information is already also summarized for treatment CTR for later discussions (Section 2.6.4).

As mentioned earlier, less than full employment should occur, since workers probably got reservation wages that employers have to guess first. In our experiment, employers hire 2.34 workers during MW1-10 on average. This is significantly different from full employment (one sample t test, $p < .001$). However, there are two hints that employers get some experience in aspired wages: The average of hired workers not only increases from 2.11 to 2.58 from MW1-5 to MW6-10 (dependent sample t test, $p < .001$) which might be attributed to wage increases because of

	treat. MW (34 obs.)		treat. CTR (30 obs.)	
	MW1-5	MW6-10	CTR1-5	CTR6-10
hrd	2.11	2.58	2.07	2.50
Pay_P	201.16	209.87	198.32	234.32
Pay_L	39.36	66.42	39.63	47.61
Pay_M	79.45	98.12	72.92	85.93
Pay_H	93.56	128.53	103.80	139.47
WF	413.53	502.94	414.67	507.33

Table 2.1: Hiring frequencies and payoffs

the minimum wage m, but also during the 5 periods before and after introducing m. While in period 1, e.g., 1.56 workers are employed, this average increases monotonically to 2.76 in period 5, before it drops to 2.09 in the first period with m and then again increases monotonically to 2.79 in the last period. These findings contradict the full employment prediction of *hypothesis I*, while confirming the alternative *hypotheses II and III.* Payoff results are not that interesting, since they include payoffs of both groups: employed and unemployed workers. However, it is interesting that payoffs not only significantly increase for all workers, but also modestly for the employer, obviously due to the increase in average hiring. Welfare is also much higher in MW6-10.[35] Overall, we conclude that hiring behavior rejects the *benchmark hypothesis I*, but confirms *hypotheses IIa, IIb and III.*

Result 1*: In treatment MW, full employment does not occur on average, but hiring averages rise after the introduction of the minimum wage and during MW1-5 and MW6-10. Payoffs of all workers largely increase after the introduction of the minimum wage.*

2.6.2 The minimum wage treatment: Wages

In the experiment, contract offers do not always lead to employment and we expect them to follow a trial-and-error-path, since principals have to learn their workers'

[35]Resulting p-values of dependent sample t tests for these five variables are: $p = .399$ for Pay_P, $p = .001$ for Pay_M and Pay_H, and $p < .001$ for Pay_L and WF.

preferences. We will thus start by limiting our analysis to those offers that are accepted. These observations seem most appropriate to characterize results, since they represent factual contracted wages and exclude (extreme) trial balloon wage offers rejected by workers. Let us, from now on, synonymously use the terms *successful wage offers* or simply *paid wages* when referring to the wage offers that are accepted by workers and thus actually apply. We will discuss the negligible differences between paid wages and all wage offers at the end of this section.

To guarantee comparability between groups, we did not introduce one concrete minimum wage for all groups, but rather used a group-specific value depending on the wage offers in MW5. Therefore, the absolute paid wages alone are still only of limited explanatory power and we additionally compute the relative change (in percent), named *chg*, between the paid wage offers in MW1-5 and those of MW6-10 for each group. Table 2.2 gives averages of paid wages (and successful wage thresholds already) for all three workers in MW1-5 and MW6-10 as well as the average minimum wage. In the fifth column the average of relative changes is given. The last column gives the number of observations for MW1-5, MW6-10 and *chg*.[36]

	MW1-5	m	MW6-10	chg	# of obs.
w_L	57.20	69.03	73.35	29.37 %	34, 34, 34
w_M	99.65	–	111.40	14.44 %	34, 34, 34
w_H	149.71	–	161.75	10.93 %	33, 33, 32
t_L	43.08	69.03	70.34	104.65 %	34, 34, 34
t_M	83.84	–	100.12	28.96 %	34, 34, 34
t_H	136.50	–	150.34	15.11 %	33, 33, 32

Table 2.2: Treatment MW, paid wages and successful wage thresholds

Averages of paid wages in MW1-5 are 57.20 for workers L, 99.65 for workers M and 149.71 for workers H, which obviously rejects the *benchmark hypothesis*, but also all other hypotheses, since it is significantly more than 40 percent of the

[36] In one group, worker H was never hired in MW1-5 and in another group, H was never hired in MW6-10. Thus, we have 33 observations there and can compute relative changes in 32 cases.

respective pies.[37] Instead, paid wages amount to about 50 percent of the respective pies for workers M and H and to a little larger fraction for workers L.

Result 2: *In the first five periods of MW, workers earn more than 40 percent of the respective revenues* R_i*.*

Paid wage averages increase from MW1-5 to MW6-10 to 73.35 for L, 111.40 for M and 161.75 for H. The average relative wage change is highest for workers L who are the only ones immediately affected by the minimum wage. Their wages increase by 29.37 percent on average.

Our main research hypothesis predicted that minimum wage spillovers occur, i.e., that wages of the medium and the high productive worker also increase in MW6-10, although these workers already earned more than the minimum wage in p5 (and MW1-5). Since $w_L \leq m \leq w_M \leq w_H$ always holds (see the sections before), the wages of workers M and H, indeed, do not have to increase, but the relative wage changes for these groups suggest that they nevertheless do, i.e., they suggest that minimum wage spillovers exist. The relative increase for workers M is 14.44 percent and thus still half as high as the increase for workers L. The increase for high productive workers is a little lower with 10.93 percent. We can test whether the increases are statistically significant by performing dependent sample t tests that use the averages of MW1-5 and MW6-10 for each group. Since we hypothesized wage increases for all workers, we perform the tests one-sided. All tests deliver significant results ($p < .001$ for workers L and M and $p = .010$ for workers H).

Before summarizing results, let us further analyze the relation of minimum wages and worker's wages. The average minimum wage is $m = 69.03$ and the average of paid wages to workers L in MW6-10 is only a little, yet significantly higher than this minimum wage ($w_L = 73.35$). This means that there is some reason to believe that a *fairness perception effect*, as FFZ2006 found it, exists. We calculated the relative change between the wages L in MW6-10 and the minimum wage for each group. On average, paid wages are about 6.5 percent higher than the minimum wage[38] which is only a little less than the about 8 percent FFZ2006 found. To put

[37] One sample t tests against 40 (worker L), 80 (worker M) and 120 (worker H), respectively. Resulting p-values are: $p < .001$ in all three cases.

[38] Dependent sample t test, two-sided: $p < .001$.

it another way: the difference between the successful wage offers to workers L in MW1-5 and the minimum wage amounts to about 23 % and is thus responsible for three fourths of the 29.5%- wage increase we observe.

The question that now arises is: Is it likely that similar effects are accountable for the wage increases we observed for workers M and H, i.e., is there reason to believe that the minimum wage spillovers should be attributed to fairness perception effects instead of *relative income effects*?

The wages paid in MW1-5 to workers M and H are on average already about 45.5 percent and 119 percent, respectively, higher than the minimum wage, i.e., the minimum wage cuts off a part of the interval of feasible wages much below the wages paid to M and H. Considering these large differences, the intuitive intermediate conclusion should be that there is little reason, if any at all, to believe that fairness perception effects similar to that for workers L should be of importance.

Although there is no one-to-one experimental evidence for this intuitive reasoning, the study of Falk and Fischbacher (2006) gives some justification. In their experiment, recipients had to evaluate the kindness of hypothetical splits of 10 Swiss Francs offered by a proposer. The kindness had to be rated on a scale ranging between 0 (very unkind) and 100 (very kind). In scenario (i), participants were told that every integer split was possible, i.e., 10 for the proposer and 0 for them ($(10, 0)$), or one unit for them ($(9, 1)$), or $(8, 2)$ and so on until $(0, 10)$. On the other hand, in scenario (vii), only the offers $(2, 8)$, $(1, 9)$, and $(0, 10)$ were feasible, i.e., the lowest offers of the original scenario were disallowed. We think that comparing these two scenarios resembles our experimental setting quite well: scenario (i) is similar to the first five periods without the minimum wage (MW1-5) and scenario (vii) for the later five periods (MW6-10) in which the minimum wage restricts the feasible offers.

Unsurprisingly, the *kindness evaluations* of the offer $(2, 8)$ in scenarios (i) and (vii) largely differed according to Falk and Fischbacher (2006). Responders evaluated kindness of $(2, 8)$ with +62.0 in (i) and only with +40.8 in (vii) where $(2, 8)$ was the lowest feasible offer. Evaluations were still differing a little for $(1, 9)$ with +68.0 in (i) and +62.0 in (vii). Note that the value for $(2, 8)$ in (i) is exactly equal to the value for $(1, 9)$ in (vii) which means that a larger amount had to be offered to achieve the same level of perceived kindness. Roughly transferred to our setting this can be used to explain the fairness perception effect for workers L: the new lower bound of $(2, 8)$, resembling our minimum wage, is obviously perceived as being unkind and might thus be avoided by mindful principals.

However, our observations for M and H are much different: Their wages are much larger than the minimum wage in MW1-5 and MW6-10 and are thus much more comparable to the extreme offer of $(0, 10)$ in scenarios (i) and (vii). In fact, the kindness evaluations of $(0, 10)$ the authors reported were almost identical. Average kindness amounted to +72.3 in (i) and was even a little higher with +73.4 in (vii). These results strengthen our intuitive argument that fairness perception effects for workers M and H should be quite modest, if existing at all.

After all, the considerations just presented do, at least, not disqualify the relative income effect as the main source of the positive minimum wage spillovers we observe. However, it will not be until the discussion of the control treatment, before we can give a more definitive answer. For now, we limit ourselves to summarize that the predictions of *hypothesis III* are essentially confirmed.

Result 3: *Paid wages in MW1-5 and MW6-10 are highest for workers H, followed by paid wages for workers M and L. In MW6-10, paid wages are higher than in MW1-5 for all workers, i.e., there are positive minimum wage spillovers for workers M and H.*

We now want to visualize and deepen the preceding analysis by discussing relative wage changes in all 10 periods. For this purpose, we introduce $rel_{i,p}$ as a measure of relative wage change for worker i in period p. We define it as the fraction between the wage in period p (w_p) and the average of paid wages in MW1-5 (w_{1-5}) times 100 minus 100, i.e., this variable gives the relative change in percent between the wage paid in a period compared to the average wage before the introduction of the minimum wage. Figure 2.2 depicts all ten periods on the horizontal axis and gives the averages of $rel_{i,p}$ on the vertical axis.

The relative changes before the introduction of the minimum wage are quite moderate for all workers and lie mainly between -5 % and +5 %. As discussed before, wages increase during MW1-5 which is reflected in Figure 2.2 by the positive slopes of the hypothetical lines connecting $rel_{i,1}$ and $rel_{i,5}$. The minimum wage forces wages for L to largely increase in MW6-10. Consequentially, from period 6

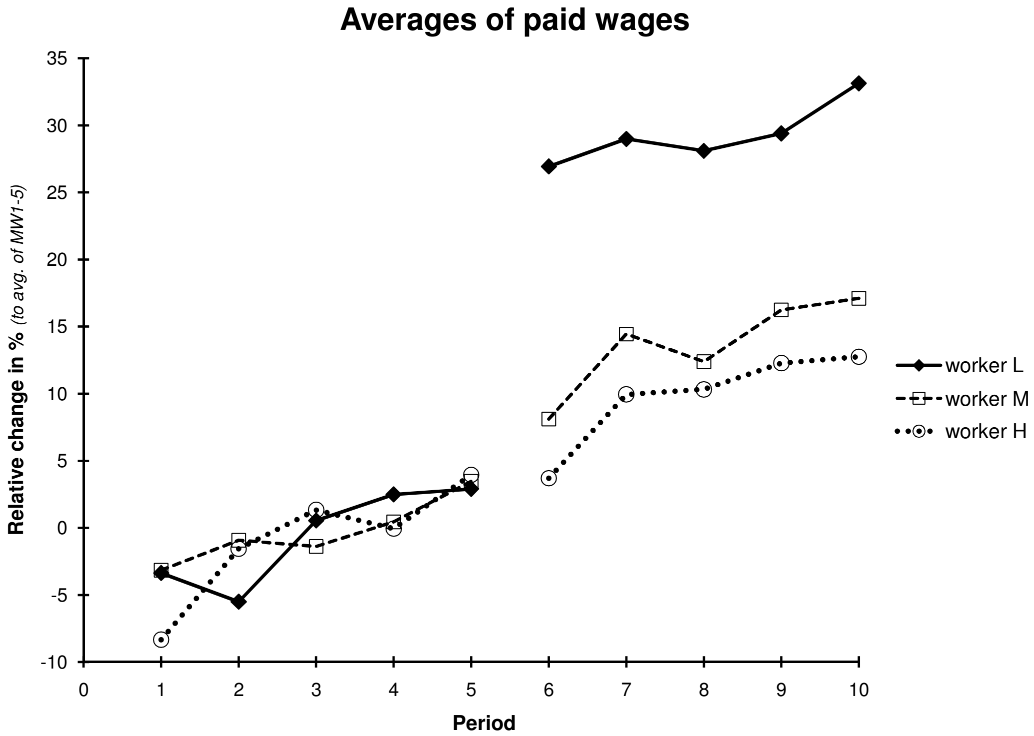

Figure 2.2: Treatment MW, relative changes in paid wages

on the relative change for workers L lies between +27 % and +33 % and is thus much higher than during periods 1 and 5. Except for a little bump from periods 7 to 8, relative changes increase for later periods again, i.e., the average of paid wages is higher at the end of the experiment than shortly after introducing the minimum wage. This trend for what is called *seniority wages* in literature was already visible on the more aggregate level in Figure 2.1. It could be a result of the higher hiring figures discussed in Section 2.6.1 or *adaptation effects* could play a role, i.e., workers' aspired wages might have continually increased (see Section 2.2.2) and employers might have anticipated this. We will pick up this thought when discussing threshold wages below.

Relative wage changes for the medium productive workers M are higher in each of the periods 6 to 10 (between about +8 and +17 %) than in each of the periods 1 to 5. There is a small jump from the last period without minimum wage to the first afterwards, i.e., the minimum wage spillover is clearly visible. Again, there is an increase in wage offers from period 6 to 10 (with a bump at period 8).

There are two minor differences for workers H: a) the wage increases in MW6-10 are a little lower than for workers M and range between +4 and +13 % and b) paid wages, and thus relative changes, are almost equal in periods 5 and 6, before they quite sharply increase in period 7 and afterwards.

Now, to conclude this subsection let us quickly broaden the view to all wage offers (those that were accepted and those that were rejected) instead of focusing on paid wages as before. Table 2.3 gives these data (again for thresholds as well).

	MW1-5	m	MW6-10	chg	# of obs.
w_L	53.35	69.03	72.94	38.46 %	34, 34, 34
w_M	93.86	–	109.88	19.93 %	34, 34, 34
w_H	137.79	–	158.91	19.56 %	34, 34, 34
t_L	47.31	69.03	70.54	88.43 %	34, 34, 34
t_M	86.96	–	102.20	22.98 %	34, 34, 34
t_H	147.32	–	154.31	8.07 %	34, 34, 34

Table 2.3: Treatment MW, all wage offers and wage thresholds

There are only minor differences to paid wages given in Table 2.2: The average over all wage offers is, of course, a little lower than in the subgroup of eventually successful wage offers in MW1-5 (about 6 to 9 percent), but this difference almost completely vanishes in MW6-10, which is not surprising given the increased hiring figures mentioned before. Consequentially, the percentages of the relative changes are a little higher this time. The increase for workers L is about 38.5 percent and again twice as high as for workers M (about 20 percent). However, the increase for workers H is only marginally lower now with about 19.5 percent. Performing the same tests as above for paid wages, i.e., checking benchmark levels, wage spillovers and the increase above the minimum wage, gives qualitatively identical results.[39]

[39] Tests results for the one sample t tests again the 40 %-prediction are: worker L: $p < .001$; worker M: $p < .001$; worker H: $p = .009$. One-sided dependent sample t tests comparing the wage average in MW1-5 with that in MW6-10 give us $p < .001$ for all workers. The p-value of the dependent sample t test comparing the wage of worker L in MW6-10 with the minimum is $p < .001$.

2.6.3 The minimum wage treatment: Threshold wages

Averages for workers wage thresholds were already given in the tables before. Since thresholds do not directly determine the contracted wages when they are lower than offers, we think it is more interesting to focus on all wage thresholds (Table 2.3), rather than limiting the view to those that lead to employment. Again qualitative differences between these two data sets are negligible and will be given in footnotes occasionally.

First of all, notice that thresholds in MW1-5 follow the same order as wages. According to Table 2.3 they are lowest for the low productive worker with $t_l = 47.31$, followed by those for the medium productive worker, $t_M = 86.96$, and are highest for the most productive one, $t_H = 147.32$. The fractions of the respective pies demanded thus lie between 40 and 50 percent. Just like for wages, the average minimum wage lies between the thresholds for L and M. After the introduction of the minimum wage in MW6-10 the thresholds increase to $t_l = 70.54$, $t_M = 102.20$ and $t_H = 154.31$, respectively. One-sided dependent sample t tests tell us that the increases for L and M are highly significant ($p < .001$) and the increase for workers H is significant on the 5%-level ($p \approx .050$).[40] This suggests that the minimum wage spillovers we observe are an echo of workers' rising wage demands and are not just founded in wrong assessments of workers' behavior by principals.[41]

We also calculated averages of relative changes again. The thresholds increase by 88.43 percent for workers L, 22.98 percent for M and 8.07 percent for H, i.e., the order is qualitatively the same, yet more pronounced than for (paid) wages. We visually present these findings in Figure 2.3 that is constructed analogically to Figure 2.2.

Except for a peak in period 2 and modest decreases in periods 5 and 10, thresholds are rather stable during MW1-5 and MW6-10. The spillover effects are clearly visible. Since thresholds do not increase during MW1-5 and MW6-10, respectively, *adaptation effects* on behalf of the workers as a cause for the seniority wage structure observed for wages are unlikely.

As for paid wages before, the figure suggests that relative changes for workers

[40] The more exact p-value is $p = .0502$.

[41] The results are similar for the sub-sample of thresholds that lead to employment. Here, p-values are $p < .001$ for L and M and $p = .002$ for workers H (same tests as before).

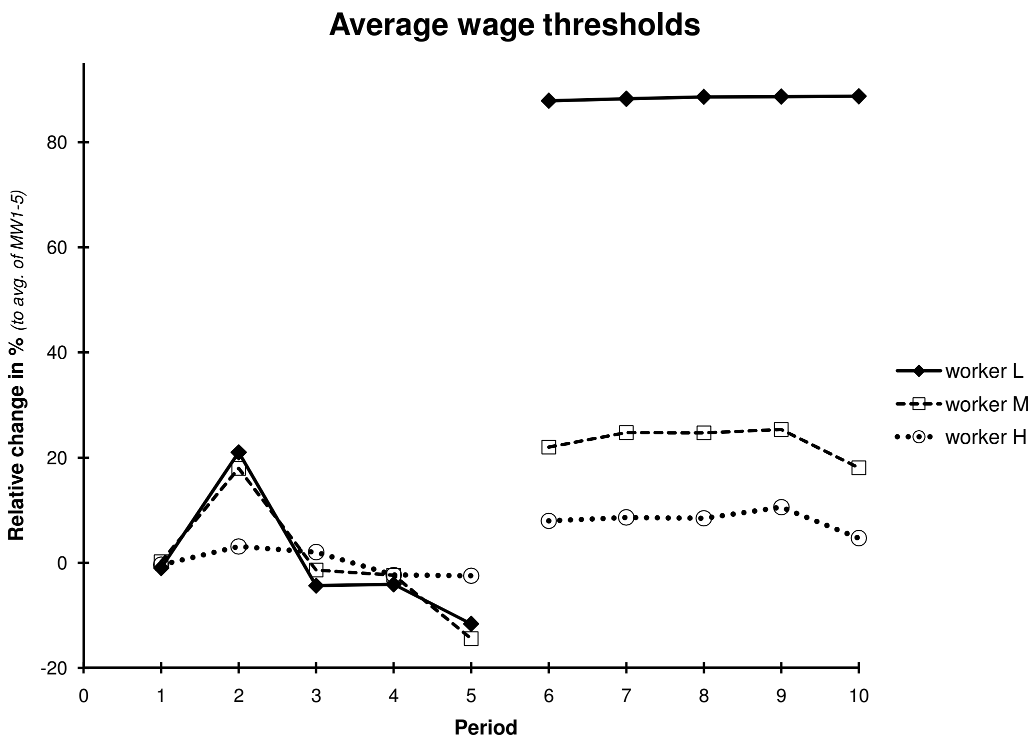

Figure 2.3: Treatment MW, relative changes in wage thresholds

M are larger than for workers H. Due to the dependencies in our data, there is no statistically sound method to test this last difference. However, we think it is fair enough to summarize that the comparisons of relative wage and threshold changes give some support to our own and other authors' theoretical and empirical conjectures that the relative income effect is more important for those workers whose earnings are (comparatively) nearer to the minimum wage (workers M) than for workers whose wages are higher (workers H).

Comparing the average wage thresholds for workers L to the minimum wage is also interesting. The minimum wage is on overage 85.45 percent higher than the threshold wages during MW1-5. Additionally, we calculate the relative change between the minimum wage and the threshold wages in MW6-10 for all 34 groups. On average wage thresholds are only about 2 percent higher than the minimum wage, which is also reflected in the averages given in Table 2.3 ($m = 69.03$ and $t_L = 70.54$). This is even more interesting when keeping in mind that we explicitly disallowed thresholds to be lower than the minimum wage to prevent them from being meaningless. Since FFZ2006 did allow workers to name threshold wages lower than the minimum wage, the authors did not compute the relative change and we cannot compare our results to theirs. Even the small differences between reservation and minimum wages we find are significant using a dependent sample t test ($p < .001$). Nonetheless, larger differences between demanded and minimum wages could have been expected considering the experimental results of McCabe et al. (2003) or Falk and Fischbacher (2006) discussed in Sections 2.2 and 2.6.2.

In summary, the fairness perception effect exists for thresholds, but is of little magnitude only. Apart from this, *Hypothesis III* is confirmed again.

Result 4: *Wage thresholds in MW1-5 and MW6-10 are highest for workers H, followed by those for workers M and L. In MW6-10, thresholds increase for all workers, and workers L demand only a little more than the minimum wage.*

We now want to finish discussing treatment MW by comparing the wage thresholds with the revenue relations. Let us define

$$tp_{i,j} = \frac{t_i/w_j}{R_i/R_j} \qquad \text{and} \qquad tp_{i,-i} = \frac{t_i/w_{-i}}{R_i/R_{-i}}$$

with $i, j \in \{L, M, H\}$, $i \neq j$, and w_{-i} (R_{-i}) giving the average of wages (revenues) of all workers except worker i. For $tp_{i,j}$ (or $tp_{i,-i}$) values larger than 1 indicate that worker i's threshold wage in comparison to worker j's wage (or all other workers' wages) is larger than the respective revenue difference, i.e., values larger 1 mean that worker i demands *"more than he deserves"* in comparison to the revenue differences. Values smaller 1 indicate the opposite. In economic and psychology literature (see Frank (1984a) for a brief introduction) it is often argued that people tend to demand more for themselves than objective criteria would allow. Table 2.4 gives all averages $tp_{i,j}$ and $tp_{i,-i}$ for all workers i. The second row defines for which reference worker (group) the value tp is calculated.

Before the introduction of the minimum wage, workers L and H, indeed, demand more than their revenue fractions. For workers L, for example, the divergence is $tp_{L,H} = 1.17$, i.e., his wage threshold in relation to the standing offer to H is 17 percent larger than the revenue relation. A two-sided one sample t test shows that

	MW1-5				MW6-10			
worker (group) j	L	M	H	$-i$	L	M	H	$-i$
$tp_{L,j}$	–	1.08	1.17	1.12	–	1.31	1.40	1.35
$tp_{M,j}$	.89	–	1.09	.99	.71	–	1.00	.89
$tp_{H,j}$	1.27	1.38	–	1.32	.72	.94	–	.85

Table 2.4: Treatment MW, wage thresholds and revenue relations

this is significantly different from 1 ($p = .044$), while the modestly lower value for his demands in comparisons to both other workers, $tp_{L,-i} = 1.12$, becomes insignificant ($p = .086$, same test).

The values for workers H are a little misleading, since they are heavily influenced by one extreme outlier with $tp_H > 10$ that was more than six times larger than the second-highest value. Without this group the values are only a little different from 1: $tp_{H,L} = .99$, $tp_{H,M} = 1.11$, and $tp_{H,-i} = 1.05$. But even when not excluding the outlier, the differences between all tp_H and 1 given in Table 2.4 are statistically insignificant due to the large variances.[42]

Workers M almost exactly demand what they deserve in comparison to both other workers in MW1-5: for them $tp_{M,-i} = .99$ is, of course, insignificantly different from 1 (two-sided one sample t test: $p = .856$). It is interesting, however, to note that this aggregate result conveys the rather large differences between their demands with respect to the wages of workers L and H. While medium productive workers demand more than they deserve in comparison to workers H, $tp_{M,H} = 1.09$ (not differing from 1, same test, $p = .282$), they want less than they deserve with regard to worker L, $tp_{M,L} = .89$ (almost significantly different from 1, same test, $p = .061$). This suggests that for workers M the wage of worker L is insofar less important as he is willing to get underpaid in comparison to him as long as he is overpaid with regard to worker H.

In MW6-10 the observations for workers L are not that meaningful due to the minimum wage. But here, a similar effect as for workers M now occurs for M and H. For workers M, the wage relation to worker H, $tp_{M,H} = 1.00$ is much larger

[42] One sample t test as before, p-values are: for $t_{H,L}$: $p = .334$; for $t_{H,M}$: $p = .175$; for $t_{H,-i}$: $p = .252$.

than that to worker L, $tp_{M,L} = .71$ and the high productive workers are now much more willing to accept an underpayment with regard to worker L than to worker M ($tp_{H,L} = .72$ and $tp_{H,M} = .94$) which could have been expected since the medium productive worker is most similar to him.

2.6.4 The control treatment

We reported minimum wage spillovers for workers M and H. However, for these workers we did not only find that the wage averages of MW6-10 were larger than that of MW1-5, but we also presented visual evidence – in Figures 2.1 and 2.2 – that wages increased during MW1-5 and MW6-10. This might leave the reader with the question whether the wage spillovers may not originate in the introduction of the minimum wage that triggers *relative income effects* as our theoretical model argued, but are rather an artefact due to a general tendency for increasing wages in experiments like ours. Similar trends did not occur for wage thresholds. Nevertheless, we did a control treatment CTR in which the first five periods (CTR1-5) were played exactly as in treatment MW, but then no minimum wage was introduced and the same five periods were played again (CTR6-10).

If the minimum wage is not triggering relative income effects that cause the minimum wage spillovers, we should observe equally high wage increases from CTR1-5 to CTR6-10 as we did before for MW1-5 and MW6-10. On the contrary, if wage spillovers are due to the minimum wage, the increases should be higher in MW than in CTR. We thus hypothesize in this section that wage increases in treatment CTR are lower than those in treatment MW.

The hiring and payoff averages previously given in Table 2.1 confirm the expectation that the first five periods of MW and CTR do not differ that much. For example, hiring averages are 2.11 in treatment MW and 2.07 in treatment CTR. When additionally splitting the hiring averages to all three workers, we find that a little more workers M are hired in MW than in CTR (.79 in MW and .71 in CTR), but less of type H (.62 in MW and .69 in CTR). Repercussions of this are visible in payoffs, but none of this slight differences is significant according to two-sided two sample Welch-Satterthwaite t tests (*WS tests*).[43] The same holds for paid wages:

[43]The p-values for all variables in Table 2.1 range from $p = .404$ for Pay_H up to $p = .968$ for

in CTR1-5 they are $w_L = 58.89$, $w_M = 103.36$, and $w_H = 154.35$, in MW1-5 they are $w_L = 57.20$, $w_M = 99.65$, and $w_H = 149.71$. These differences are statistically insignificant.[44]

Let us now briefly describe results for CTR6-10 and their differences to those in CTR1-5. Table 2.5 presents the same data for paid wages in CTR as Table 2.2 did before for MW with the exception that the minimum wage column is, of course, left out. Instead a last column is added that gives the relative changes for MW, again.

	CTR				MW
	CTR1-5	CTR6-10	*chg*	# of obs.	*chg*
w_L	58.89	57.92	9.64 %	29, 29, 29	29.37 %
w_M	103.36	106.54	7.91 %	28, 29, 28	14.44 %
w_H	154.35	162.24	3.08 %	28, 30, 28	10.93 %

Table 2.5: Treatment CTR, paid wages

First of all, notice that although the paid wages for workers L slightly decrease from $w_L = 58.89$ in CTR1-5 to $w_L = 57.92$ in CTR6-10 the relative wage change is positive with 9.64 percent. This is due to one extreme outlier with an increase of more than 300 percent. Since differences between CTR and MW are large anyhow and not in our main interest, we do not bother detailing what would change if we excluded this outlier. Furthermore, a one-sided Wilcoxon paired sample test refutes the hypothesis that wages increase for workers L from CTR1-5 to CTR6-10 ($p = .467$).[45]

The medium productive workers' average wages are 103.36 in CTR1-5 and increase to 106.54 in CTR6-10. This increase is significant according to a one-sided Wilcoxon paired sample test ($p = .025$). For workers H the increase from 154.35 in CTR1-5 to 162.24 in CTR6-10 is slightly insignificant using the same test ($p = .068$). These intermediate results suggest that wages for M and H are also increasing in CTR.

WF. Data on request.

[44] Mann-Whitney U-tests, p-values are: $p = .829$ for w_L, $p = .409$ for w_M, and $p = .694$ for w_H. Results are similar when looking at all wage offers instead. Data on request.

[45] We choose the one-sided hypothesis here and below, since we did so for MW before.

Are these wage increases in CTR at least smaller than in MW? For answering the question, we start by discussing differences for workers L that are, of course, rather extreme due to the influence of the minimum wage in treatment MW. It thus suffices to look at the aggregate averages of relative wage changes given in Table 2.5 that are 29.37% for MW and only 9.64% for CTR. The hypothesis that both changes are equal is refuted in favor of the alternative that wage increases are larger in MW using a one-sided two sample Mann-Whitney U-test (*U-test*) ($p < .001$).

Visually depicting the relative changes in paid wages in treatments CTR and MW for workers M and H is helpful. This is done in the upper half of the following Figure 2.4. The lower half of Figure 2.4 deals with wage thresholds and will be discussed a few paragraphs later.

Relative wage changes in MW and CTR for workers M and H are nearly indistinguishable in periods 1 to 5, i.e., they vary between -5 % and +5 % and are increasing over time, although the latter trend is less pronounced for workers M in CTR than in MW. However, for both workers the differences between treatments MW and CTR are clearly visible in periods 6 to 10.

In each of the periods 6 to 10, the relative wage change for workers M in treatment MW is more than twice as high as in treatment CTR. Furthermore, the wage increase during CTR6-10 from about +4 % to +7.5 % is a little less pronounced than that during MW6-10 from about +8 % to about +17 %. On aggregate level the wages of workers M increase by 14.44 % in treatment MW, but only for 7.91 % in the control treatment. This difference is significant according to a one-sided U-test ($p = .012$).

In period 6, relative wage changes for workers H are only a little higher in MW (+3.70 %) than in CTR (+2.03 %). Afterwards wages and thus relative changes largely increase in MW, but eventually decrease a little in CTR so that changes are at least 3 times as high in MW in each of the periods from 7 to 10. Since visual inspection suggested differing variances we perform a Mann-Whitney U-test and, additionally, a robust rank-order test *(U-test (rro test))*. They tell us that the difference between relative wage change averages in MW (10.93 %) and CTR (3.08 %) is only slightly insignificant with a p-value of $p = .055$ (.054).

Overall, these results suggest that the wage spillovers are larger in the minimum

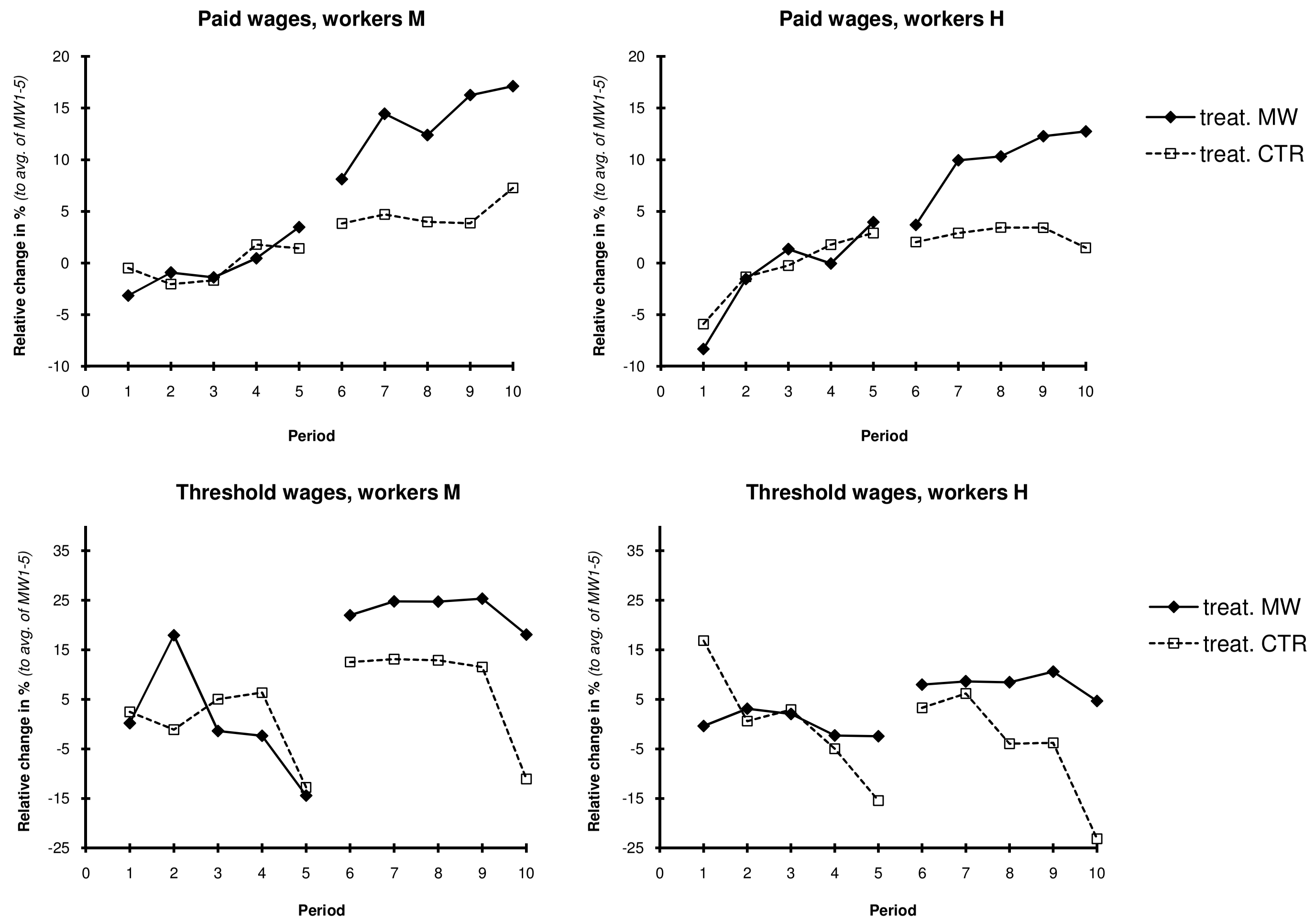

Figure 2.4: Treatments MW and CTR, relative changes in paid wages and wage thresholds

wage treatment MW than in treatment CTR. This implies that the wage spillovers in treatment MW originate in the introduction of the minimum wage that triggers relative income effects, and are not just an artefact caused by a tendency for wages to increase over time. The results do not differ much when broadening the view to all wage offers.[46]

Let us conclude this section by comparing differences in threshold wages in CTR and MW. Table 2.6 gives averages for the whole group of all wage thresholds for CTR just as Table 2.3 did for MW. Additionally, changes for MW are given in the last column.

	CTR				MW
	CTR1-5	CTR6-10	*chg*	# of obs.	*chg*
t_L	52.77	52.86	1.76 %	30, 30, 30	88.43 %
t_M	95.75	101.05	7.80 %	30, 30, 30	22.98 %
t_H	140.83	137.50	-4.27 %	30, 30, 30	8.07 %

Table 2.6: Treatment CTR, wage thresholds

Trends are even clearer than for paid wages: Threshold wages for workers L are almost identical in CTR1-5 and CTR6-10 with an average increase of only 1.76 %, while due to the minimum wage they inevitably largely increase in MW (88.43 %).

The moderate increase of 7.80 % for thresholds of workers M in CTR is much smaller than the 22.98 %-increase in MW. This difference is significant according to a one-sided WS test ($p = .003$). Threshold wage averages for workers H decrease from CTR1-5 to CTR6-10 for -4.27 % and are thus lower than the already only modest increase of 8.07 % in treatment MW. This difference is also significant using the same test again ($p = .016$).

[46] Relative wage changes for workers L (M) are 38.46 % (19.93 %) in MW, but only 9.99 % (11.20 %) in CTR; one-sided WS test: $p < .001$ (p = .064). At first glance, the difference for workers H is rather small with 19.56 % in MW and 16.13 % in CTR. However, the high relative wage change for CTR is mainly due to an extreme outlier case in which the principal offers only 42, on average, to his worker H in CTR1-5 and thus never employs him, but offers 148 in CTR6-10. The relative wage change in this group is almost 6 times as high as the second highest relative change, while the analog factor is only 1.6 in MW. When excluding the outlier, the average relative change drops from 16.13 % to 7.98 % and the difference to the 19.56 %-increase in MW is significant (one-sided U-test: $p = .024$).

These results suggest that not only wage spillovers in MW are due to the minimum wage and relative income effects, but this minimum wage also changes workers threshold wages much more lasting than time does.

Figure 2.4 visualizes these insights for workers M and H and since results are qualitatively unchanged when restricting attention to successful wage thresholds only[47], we conclude:

Result 5*: In p1-5, results for CTR and MW are almost identical. Increases in wages and threshold wages are larger in MW than in CTR.*

2.7 Conclusion

Minimum wages are a labor market institution regularly discussed and implemented across the world. While the public debate often centers around questions of the proportionateness of a society's wage distribution and dangers of financial and social exclusion of people with low wages (see, e.g, The Australian (2009), Irish Times (2009a), or Washington Post (2009)), scientific studies, both theoretical and empirical, predominantly focus on direct fiscal and employment effects of minimum wages. Our study does not investigate these fiscal and employment effects. Instead, it analyzes the indirect effects of minimum wages, particularly the repercussions of minimum wages on the wages of workers who earned more than this minimum wage before its introduction. In cases in which wages of those workers increase after introducing a minimum wage, literature speaks of a minimum wage spillover. The question whether such minimum wage spillovers exist or not should be important for lawmakers and their advisors, not only when considering introducing a minimum wage, but also when raising an existing minimum wage.

We are confident that we enriched the knowledge about minimum wages by a) theoretically analyzing minimum wage spillovers in a model that focuses on relative income preferences and heterogeneously qualified workers, and by b) experimentally testing our theoretical predictions.

We started by evaluating the existing literature and singled out the best es-

[47] The p-values of one-sided U-tests (and rro tests for workers L and H) are $p < .001 (< .001)$ for L, $p = .004$ for M, and $p = .043 (.042)$ for H.

tablished cause for minimum wage spillovers, namely peoples' preferences for a favorable relative income position. Two concrete classes of utility functions that capture relative income preferences were proposed in literature. In our relative income model, we derived main results without committing to one of these specific classes of functions and were able to show that a set of rather general assumptions about relative income preferences suffices to qualitatively analyze minimum wage spillovers. The model predicted that minimum wage spillovers occur and suggested that they originate in workers' relative income preferences.

Our experimental design followed the relative income model and thereby excluded some of the other causes for minimum wage spillovers discussed in theoretical and empirical literature like substitution effects, for example. The experiment is thus, on principle, not only capable of testing for minimum wage spillovers, but also of attributing them with high certainty to one specific cause, namely relative income preferences.

Our experimental results essentially confirmed the theoretical predictions: minimum wage spillovers occurred and we presented evidence that they originated in relative income preferences.

We are aware that by concentrating on relative income preferences only, one might overemphasize the magnitude of minimum wage spillovers in comparison to a real world environment in which other effects also might play a role. However, our findings do not suggest that minimum wage spillovers are only statistically significant, but of no practical relevance. On the contrary, we observe that the increases in wages of workers who earned more than the minimum wage before its introduction, i.e, the minimum wage spillovers, are almost half as high as the increases for workers whose wages have to rise, since they earned less than the minimum wage before its introduction. This clearly indicates rather strong minimum wage spillovers and we conclude that they should be considered by lawmakers when introducing or modifying minimum wage laws.

3 Employment protection and bullying

3.1 Introduction

In this chapter, we experimentally examine the relation between employment termination laws and sanctions in the workplace (*"bullying"*). The idea that under strict termination laws, bullying might be used by employers to provoke unproductive workers to quit voluntarily, has been introduced by Wasmer (2006) in a theoretical and empirical study. Even if bullying is costly, it can be a part of the employers' equilibrium strategy according to his model. In the empirical part of his work, Wasmer finds supporting evidence for the hypothesis that firms bully workers under strict employment protection laws (EPL). In particular, he reports positive correlations between several dimensions of workers' stress (as a proxy for bullying) and employment protection: stress levels increase with stricter EPL, i.e., workers are more stressed when EPL makes it harder for firms to fire their workers. In addition, the consumption of anti-depressants also increases with stricter EPL.

Instead of using field data as a proxy for bullying, we run a laboratory experiment in which we allow the employers to sanction their workers. We interpret sanctions as bullying and investigate whether or not an institutional change in the degree of employment protection has an impact on the usage of such sanctions.

In one group of treatments, the employers are allowed to issue sanctions while in the other group sanctions are not available. In both groups, we distinguish between three stylized legal rules that represent different degrees of employment protection: In the treatments without EPL, the employer can terminate the contract at will. In the treatments with strict EPL, the employer must never terminate the labor contract. In the treatments with weak EPL, the employer may only terminate a contract if the employment relation lasts for just one period, i.e., the employer can fire an employee only during the first period.

We assume the sanctions to be costly for both the employer and the worker. This rather general approach allows us to interpret sanctions or bullying as a variety of actions, starting from casual meanness up to open hostility. Moreover, we assume the employees' effort choices to be unobservable (and, thus, non-verifiable). We introduce this uncertainty since we perceive it as crucial for EPL to be meaningful at all: Under perfect monitoring, there is little room for opportunistic behavior on either side.

By examining the relation between EPL and sanctions in a laboratory experiment, we link the work of Wasmer (2006) and the numerous works on sanctions (not necessarily at the workplace) in experimental economics. Fehr and Gächter (2000a, 2002) have examined sanctions in the context of public-good experiments. Closer to the workplace context are the papers by Andreoni, Harbaugh and Vesterlund (2003), subsequently denoted as AHW2003, who have looked at a proposer-responder game, and Kirstein (2008), who examined a principal-agent setting with real effort. All these papers presented evidence that people are willing to sanction others perceived as defectors even if such sanctions are costly and the interaction is anonymous and finite.[1] The authors also have found evidence that cooperation levels are significantly larger than predicted by models assuming perfect rationality.[2]

Various attempts have been made to theoretically explain these and similar findings. For example, *social-preference models* have incorporated aspects of inequality aversion, fairness, and reciprocity.[3] In a neuroeconomical experiment, de Quervain et al. (2004) found supporting evidence for the use of these models (see Fehr et al. (2005) for an overview). We lean on *social preference models* when establishing our

[1] A short overview dealing with several aspects of human altruism and sanctioning behavior is given by Fehr and Rockenbach (2004). The terms *"sanction"* and *"punishment"* are often used synonymously in this literature.

[2] Gürerk et al. (2006) even found that sanctioning institutions are preferred over non-sanctioning institutions by a vast majority of participants.

[3] The seminal paper here is Rabin (1993). Many modifications have been proposed, see, e.g., Bolton and Ockenfels (2000), Dufwenberg and Kirchsteiger (2004) or Falk and Fischbacher (2006).

behavioral hypotheses.

Employment protection legislation is a widely discussed topic, both in economic research and in the public discourse. Many of the numerous economic papers focus on the effect of EPL on employment.[4] Even though experimental analysis of labor market legislation would allow for the implementation of institutional changes in a controlled environment, this strand of literature is still in its infancy. Falk, Huffman and MacLeod (2008), henceforth FHM2008, have analyzed the relation between EPL and rewards. Their experiment is based on four treatments: no EPL vs. weak EPL is matched with availability vs. non-availability of reward.[5] Thus, the authors disregarded the relation between employment protection and sanctions (or bullying) which is the main focus of our work. Moreover, they assume that employers can perfectly monitor their workers' effort choices. As we will draw on their findings when deriving our hypotheses, a more detailed discussion of their main results is presented in Section 3.4.

In the light of the findings by Wasmer (2006), we would expect that costly sanctions (bullying) are used less often when employers can use costless firing as an alternative incentive device to discipline their workers. Transferred to our experimental setting this means that in the treatments in which firings and sanctions are admissible, we should observe that firings are used instead of costly sanctions, i.e, sanctions should decrease in comparison to treatments in which firings are not allowed. From now on, we will speak of firings and sanctions as *substitutes* in this chapter when referring to the idea that firings are used instead of sanctions when both are available.

In contrast, Rockenbach and Milinski (2006) reported a different result for an experiment in which two mechanisms to overcome social dilemma problems were available: sanctions and indirect reciprocity. They found that costly sanctions in a public-good game did not completely disappear in treatments in which it was also possible to indirectly sanction the non-cooperators of the public-good game by

[4] The net effect of EPL on employment is controversial, see, e.g., Hopenhayn and Rogerson (1993), Mortensen and Pissarides (1994) or Garibaldi (1998). EPL has also been credited for sheltering the quasi-rents from specific investments and thus fostering the accumulation of human capital (see, e.g., Schellhaaß and Nolte (1999) or Kirstein et al. (2000)). Belot et al. (2007) try to combine these models.

[5] The same experiment is also part of the paper by Falk and Huffmann (2007).

sending less money to them in a indirect reciprocity game that was also played. If the relation between firings and sanctions in our setting were similar to that between sanctions and indirect reciprocity found by Rockenbach and Milinski (2006), then we should observe that sanctions are still used when firings are also admissible, i.e., that firings and sanctions are used as *complements*.

In summary, the three most important questions our study wants to answer are: 1) Do employers use costly sanctions (*"bullying"*) more often or more severely, the stricter the employment protection is?, 2) Do employees react to higher sanctions by voluntarily quitting their job?, and 3) Are firings and sanctions used as substitutes?

Our main results are that effort levels are higher in the treatments with sanctions compared to treatments in which sanctions are not available, but sanctions do not continually increase the tighter the employment protection becomes. We also find that sanctions and firings are often used as complements, rather than as substitutes, and thus induce workers to quit not only in treatments in which firings are prohibited, but also in treatments in which firings are allowed.

In Section 3.2, we present the model and its solution. Sections 3.3 and 3.4 contain the experimental design and the hypotheses, respectively, while the results are discussed in Section 3.5. Section 3.6 concludes this chapter.

3.2 Model

Our model is also the blueprint for the experimental design. Here, we only set up the generalized model (and point out the differences between our design and the one in FHM2008). The specific parameter settings that have been used in the experiment will be presented in Section 3.3.

We are examining a labor market with m firms (the principals) and n potential workers (the agents), where each principal can employ one agent at most. There are more agents than principals ($n > m$), which implies equilibrium unemployment. The following game is played for a finite number of T periods.

The fixed component of the workers' compensation, denoted $F > 0$, is exogenously given. This is a difference to FHM2008 who set up their model with endogenous wages.

In the first out of the T periods, the interaction starts in **Stage 1**. Each principal is endowed with a budget B and decides whether she wants to employ an agent or not. All agents decide whether they want to be employed or not.

In **Stage 2**, those employers and workers who are willing to form a labor relationship are randomly matched.[6] If the number of agents who want a job exceeded the number of principals who want to hire, then some agents are involuntarily unemployed. As an outside option the unemployed agents receive an unemployment benefit, denoted $U > 0$. The payment U is assumed to be smaller than the fixed wage, F. For unemployed agents, the period ends after Stage 2.

In **Stage 3**, each employed agent privately chooses his effort level from a discrete set $\{e_1, e_2, ..., e_K\}$ consisting of effort levels e_k ($k \in \{1, ..., K\}$) with $e_1 < e_2 < ... < e_K$. We assume the costs of effort, $C(e)$, to increase in the effort level, with $C(e_1) \geq 0$.

Let Q be the set of possible output quantities q produced in the firm. A higher effort chosen by the agent shall increase the respective principal's expected output. Therefore, we assume that each effort level e_i induces a random variable X_i with a distribution function F_i. A principal's output $q \in Q$ when employing an agent who chooses effort level e_i equals the realization of the random variable X_i.

We assume that the distribution functions of the variables X_i reflect first-order stochastic dominance such that $F_1(x) \geq F_2(x) \geq ... \geq F_K(x)$ holds for all admissible values x, and $F_{k1}(x) > F_{k2}(x)$ holds for at least one x for each pair $k1, k2$ with $k1 > k2$ and $k1, k2 \in \{1, ..., K\}$. This implies $E[X_1] < E[X_2] < ... < E[X_k]$, i.e, the expected output strictly increases in the chosen effort level. Each random variable realizes once every period and applies to all matches of this period, i.e., each worker who spends effort e_i in a specific period will produce the same amount of output in this period. The distribution functions are common knowledge.

Furthermore, we assume that for each output quantity q the probability that q results is larger than zero for at least two different effort levels. Thus, the principal cannot infer the effort choices of her worker from the observed output level with certainty. This is another difference from the setup of FHM2008, who did not introduce uncertainty on the part of the principal.

[6] See Section 3.3 for details of our experimental matching. The matching in FHM2008 is not random, but is the result of a kind of oral auction.

In **Stage 4**, the random variable realizes. All players of a match learn the actual output, which is then sold. We assume each principal to be a price taker in a competitive environment with a given market price, denoted p. Finally, all principals receive information about the average output $\overline{q}$.

In **Stage 5**, we distinguish between two different situations: in one situation, the principals are not allowed to mete out sanctions; in the other, sanctions are available. If available, one unit of sanction, denoted with $s \geq 0$, decreases a worker's payoff times a factor $f > 0$. Principals do not have the option to threaten with future sanctions. This is another difference to the setup of FHM2008, who allowed their principals to promise rewards.[7]

A principal's budget, B, is no smaller than the sum of the constant labor market fixed wage F and the maximum sanction $\overline{s}$. We furthermore make the assumption that $E[X_1] \cdot p > F$, i.e., we restrict attention to situations in which hiring is profitable, even if the worker is spending minimum effort.

In the period under scrutiny, the principal's expected profit, $E[\pi]$, thus amounts to B if she is not hiring a worker, and to $B - F + p \cdot E[X_i] - s$, if she employs an agent who chooses effort level e_i and sanctions him with s. The worker's payoff, denoted ω, is either U if he is not employed, $F - C(e_i)$ if spending effort e_i and not being sanctioned, or else $F - C(e_i) - f \cdot s$.

We also assume the difference between the fixed wage and the costs from spending minimum effort to be significantly larger than the unemployment benefit, i.e., each agent has a incentive to seek employment – at least if he anticipates not to be sanctioned.

In **Stage 6** we distinguish between three different versions of employment protection legislation (EPL):

- Under **no EPL**, a principal can fire his current agent at will. Then, if still necessary, the agent decides whether to quit or not. If a contract is not terminated by either the employer or the worker, the next period starts for this principal and this agent with Stage 3 (the effort decision of the agent).

[7] This obviously is not important for the model, but may be relevant for the experiment. One might argue that using such "cheap talk"–announcements as FHM2008 did might bias the experimental results by inducing endowment effects, see Kahneman et al. (1991).

If a contract is terminated, the next period starts for both of them with the matching in Stage 1.

- With **strict EPL**, a principal cannot fire his employed agent in Stage 6. If an agent quits, however, he and his principal start in Stage 1. If the agent does not quit, the players resume in Stage 3 of the next period.
- If, under **weak EPL**, a contract between a principal and his worker lasts for just one period, then the principal can terminate it. In this case, or if the agent quits, the players continue in Stage 1 of the next period. If the relation between a principal and the agent already lasts for more than one round, then the principal is not allowed to terminate the contract. In that case, the players continue in Stage 3 of the next period (unless the agent quits, which puts the players into Stage 1 of the next period).

The subsequent periods are played according to the same pattern. The last period $\overline{T}$ ends automatically after Stage 5.

Depending on whether sanctions are available (S) or not (NS), and the prevailing EPL regime (noEPL, weakEPL, strictEPL), six different scenarios are possible. The subgame-perfect equilibria under the assumption of perfectly rational, payoff-maximizing players lead to identical results. As an example, we now verbally derive the equilibrium for the scenario without EPL and without sanctions.

At the end of the last period $\overline{T}$, each principal who employs a worker never uses costly sanctions. Agents anticipate this and, thus, choose the minimum effort level (e_1). Since we assumed the fixed wage to be larger than the outside option, the agents want to be employed. In Stage 1, this means he is seeking employment, whereas in Stage 6 this means he does not want to quit. The expected revenues are larger than the fixed wage by assumption, implying that risk-neutral principals want to hire at the beginning of period $\overline{T}$ and would not fire their current agent in Stage 6 of period $\overline{T}-1$, since every agent should spend minimum effort in the last period. They never sanction in period $\overline{T}-1$, which induces the agent to spend minimum effort. The same argument applies to periods $\overline{T}-2$ and so on.

Thus, standard game theory predicts that sanctions never occur and they are also invariant with regard to EPL. Agents always want to be employed and spend minimum effort if hired; employers always hire.

3.3 Experimental design

The experimental treatments reflect the six scenarios of the model. We distinguish between three different kinds of employment protection – no EPL (nEPL), weak EPL (wEPL), strict EPL (sEPL) – and two sanction regimes (sanctions available (S), no sanctions available (NS)). The six treatments are numbered treatment I, II, ..., VI, see Table 3.1. Our treatments IV and V are basically comparable to treatments of the experiment conducted by FHM2008, although the stages in each period of our model differ from their model.

	no EPL	weak EPL	strict EPL
S	nEPL+S (tI)	wEPL+S (tII)	sEPL+S (tIII)
NS	nEPL+NS (tIV)	wEPL+NS (tV)	sEPL+NS (tVI)

Table 3.1: Treatment overview

Each treatment was played for 10 periods that consisted of the six stages described above. We ran two sessions of each treatment and pooled the data from both sessions. In each session, 8 principals and 12 agents were active.[8]

At the beginning of each session, participants were randomly assigned to the two different roles: principal (*"employer"*) or agent (*"employee"*), which they kept throughout the experiment. We used the term *"punishment"* in the instructions. Payoffs were calculated in GE (Experimental Currency Units) during the experiment. They were exchanged into € at a rate of 1 GE = € .1 after the experiment. Workers received an additional fixed payment of 50 GE. Representative instructions are given in Appendix A.2.

[8]In the second session of treatment IV, we had too few participants and decided to run the session with 6 principals and 9 agents (the same ratio as before). We checked whether this session was different from the session with 20 participants and found no significant differences: a Wilcoxon Mann-Whitney U-test on differences in effort levels between these two sessions is insignificant on the 5%-level (p = .426). The same holds for output, average output, payoffs and output deviations with p-values between .082 and .609. Data on request.

We choose the following parameter values: $B = 12$ GE for the budget, $F = 12$ GE for the fixed wage, $U = 7$ GE for the unemployment benefit and $p = 5$ GE for the market price. Employed agents had to choose their effort level from the discrete set $\{e_1, e_2, e_3, e_2, e_5\}$. Their effort costs were set to: $C(e_1) = 0$ GE, $C(e_2) = 0.25$ GE, $C(e_3) = 0.75$ GE, $C(e_4) = 1.5$ GE, and $C(e_5) = 2.5$ GE.

The random variables X_1, X_2, X_3, X_4 and X_5 that determined the output level probabilities (in stage 4) are given by Table 3.2.

	$q_1 = 2$	$q_2 = 3$	$q_3 = 4$	$q_4 = 5$
e_1	.50	.30	.15	.05
e_2	.40	.30	.20	.10
e_3	.25	.25	.25	.25
e_4	.10	.20	.30	.40
e_5	.05	.15	.30	.50

Table 3.2: Probability functions X_1, X_2, X_3, X_4, X_5

The average output $\overline{q}$ was revealed to the principals after Stage 4 in each period.[9] If allowed by the treatment, each principal could use up to 2 GE to invest them into sanctions s. The respective sanction investment was multiplied by a factor $f = 3$ to compute the decrease of the respective workers payoff.

If labor relationships ended, it was excluded that the same principal and agent were matched in the next period again.

Participants' period earnings were accumulated and paid anonymously at the end of the experiment. Prior to the first period, participants were asked to fill in a computerized questionnaire that checked their proper understanding of the instructions. A total of 235 students, mostly in business engineering at the Universität Karlsruhe (TU), participated in the experimental sessions which were conducted at the experimental computer laboratory in Karlsruhe, in June 2008. Average earnings amounted to € 15.42 for about 100 minutes (about € 9.25 per hour).

[9] We limited this information to one position after the decimal point.

3.4 Hypotheses

The subgame-perfect equilibria derived in Section 3.2 allow us to state our **benchmark hypotheses**. Sanctions should be $s = 0$. Effort levels should be minimal (e_1) resulting in output and average output equalling approximately 2.75. Principals should always hire and earn about 13.75 GE; agents should always want to work and earn 12 GE if hired, and 7 GE otherwise.

The experimental findings on sanctioning behavior and on labor market interactions provide evidence that the benchmark hypotheses are only of limited predictive value. In those experiments, the subjects show a tendency to spend positive effort, and they are willing to issue costly sanctions, even if the subgame-perfect equilibrium predicts zero levels. Thus, the experimental literature invites us to conjecture **behavioral hypotheses**. We mainly focus on hypotheses about effort and sanctions and draw on the studies by FHM2008 and AHW2003 as the main sources for our hypotheses.

FHM2008 studied an experiment (running for 18 periods) in which in the beginning of each period, employers could offer contracts consisting of fixed wage and a desired effort level to either one specific or all workers. As soon as a worker accepted a contract offer, he and the employer were matched and the employer's offer was removed, i.e., the employer could not hire more than one worker. Workers then chose their effort levels.

Subjects played in four treatments which differed in the EPL regime (no EPL or weak EPL) and with regard to whether or not rewards were available to the principals. Strictly speaking the principals in the experiment of FHM2008 could not fire their agents, but only opt to not offer a new contract. In particular, in the no EPL treatments, employers could decide at the end of each period if they wanted to make a private offer to the same worker again or not. FHM2008 interpret the cases in which the employer did not make such private offers as firings. Although this is a difference to our setting in which employers directly choose to fire or not, we will talk of firings when describing the experiment of FHM2008, since the authors also did this and it simplifies the following descriptions. The treatments with weak EPL are loosely comparable to ours with weak EPL, i.e., firings were only possible in the first firing period. Afterwards the dismissal barrier in the experiment of FHM2008 forced employers to offer a contract to the same worker again consisting of a fixed wage at least as high as in the previous period.

FHM2008 found that average effort levels were larger than the minimum in all treatments. Effort levels were smaller in the treatment with weak EPL and no rewards than in all other treatments.

The availability of reward instruments increased the average effort levels for both, no EPL and weak EPL. Rewards were frequently used, although they were costly.

The first employment periods in treatments with weak EPL are called *probation periods* in literature and empirical works suggested that workers might spend rather much work effort during probation periods and (much) less afterwards, since employers are only able to fire during probation periods.[10] This is named a *probation period effect*.

FHM2008 found that if rewards were not available, overall efforts were much lower for weak EPL than for no EPL, with the exception of probation periods of weak EPL where efforts were slightly higher than in the no EPL treatments. If the principals were allowed to use rewards, then this effort effect almost completely vanished. In the treatment with weak EPL and without rewards, efforts in probation periods were much higher than afterwards, i.e., a *probation period effect* occurred.

AHW2003 studied a proposer-responder game that started with a proposer offering some amount of money to a responder under one restriction: The offers had to amount to at least a sixth of the total amount of money.

In one treatment, the responder could only accept this offer (as in a dictator game), whereas in three other treatments he could either sanction, reward, or sanction and reward the proposer.[11] This resembles our experiment when interpreting the proposer's offer in AHW2003 as our agent's effort choices. Amongst other things, AHW2003 found that proposers offered more than the minimum amount even if no sanctions were admissible. The availability of sanctions increased the offers. This increase came at a cost, however, since sanctions were frequently used. Rewards were more successful in increasing effort levels on an aggregate level.

[10] Ichino and Riphahn (2005) use absenteeism time as a proxy for effort in an empirical study and find similar results.

[11] AHW2003 named these options *"stick"*, *"carrot"* and *"carrot/stick"*.

According to the experimental work on sanctions and rewards,[12] we should expect effort levels to be above the minimum in our experiment, even in the treatment without sanctions. The proposed explanations include intrinsic motivation, social welfare preferences, or fairness considerations (see Andreoni (1995), Charness and Rabin (2002), Frey and Osterloh (2002)).

Hypothesis 1: *In all treatments, agents spend more than the minimum effort.*

According to FHM2008, rewards have a positive impact on effort levels in their EPL game, especially for the treatments with weak EPL. We combine this with the results of AHW2003 who found that the availability of sanctions also increases co-operation. Thus, we expect effort levels to be higher if principals may use sanctions, compared to the treatments without sanctions with the same degree of EPL.

According to FHM2008, introducing weak EPL instead of no EPL without access to rewards leads to decreased effort levels in their experiment, while this effect almost completely disappears in treatments with rewards. The same holds for the probation period effect: It is strong in the treatment with weak EPL and without rewards, but is offset in the treatment with rewards. When transferring this to our setting, the gradual weakening of the principal's strategic position from no EPL over weak EPL to strict EPL can be expected to decrease effort levels, while this effect may be weaker in treatments tI - tIII in which sanctions are possible.

Hypothesis 2: *a) Effort levels are higher if principals may use sanctions, compared to the treatments without sanctions with the same degree of EPL.*

b) In treatments without sanctions, the average effort levels are lower, the tighter the EPL. The availability of sanctions lowers this effect.

c) Probation period effects occur in treatments with weak EPL. The availability of sanctions lowers this effect.

AHW2003 and FHM2008 find that rewards or sanctions are used quite frequently, even though they are costly. Other experimental evidence supports this (see, e.g., Fehr and Gächter (2002), Kirstein (2008)). Fairness and reciprocity preferences can be the rationale for such behavior (see, e.g., Rabin (1993), Dufwenberg and Kirchsteiger (2004), Falk and Fischbacher (2006)). Within a treatment, principals should

[12] See Fehr and Schmidt (2007) or Kirstein (2008).

use more severe sanctions when suspecting that the effort spent by their agent was rather low. This would be in line with AHW2003 and FHM2008 (however, in their experiments effort is observable). In our model and experiment, the principals cannot observe effort, but output and average output. Thus, it is likely that they use output deviation, defined as the own output minus the average output, as basis for their decisions.

Hypothesis 3: *Costly sanctions are frequently used by principals. Within a treatment, the smaller the output deviation, the more severe the sanctions.*

The empirical findings of Wasmer (2006) suggest that sanctions are used more frequently (or more severely) the tighter the employment protection legislation is, whereas FHM2008 only analyzed rewards and AHW2003 did not analyze employment protection.

On the question whether firings and sanctions in our experiments are used as substitutes or complements, we hypothesize them to be substitutes. This means that employers should use (costless) firings as a disciplinary device instead of costly sanctions when both are available. Therefore, sanctions should be lower in those treatments that in the treatments in which firings are not admissible. This seems an intuitive interpretation of the findings by Wasmer (2006).

Hypothesis 4: *Sanctions occur more often (or are higher) the tighter the EPL is. Firings are used instead of sanctions when both are admissible.*

3.5 Experimental results

At first, we will give some general results dealing with hiring behavior, match lengths and the benchmark predictions (Section 3.5.1) and then discuss effort levels (Section 3.5.2). This is necessary to fully understand the results on sanctions, extensively presented in Section 3.5.3. Finally, Section 3.5.4 briefly discusses resulting payoffs.

3.5.1 General results

Agents virtually always try to be hired. In each treatment, up to 240 agents have to decide if they want to be employed or not.[13] The relative frequencies of applying agents ranges between 99.2 % and 100 %. The vast majority of employers also wants to employ, but they are a little more reluctant to hire for tighter EPL. With sanctions available, the concrete relative frequencies are 99.4 % with no EPL (tI), 96.3 % with weak EPL (tII), and 94.4 % with strict EPL (tIII). Values are almost equal without sanctions: 99.3 % in tIV, 98.1 % in tV, and 93.8 % in tVI. Overall these results largely confirm that our setting almost induced maximum employment as expected.

Let us now describe labor fluctuation. For this reason we first count the cases in which an employer does not hire the same worker in period $t + 1$ as in t (either because she fires the worker, she switches from hiring to non-hiring or vice versa, or the worker quits). Then, we relate the number of this cases to the hypothetical maximum, i.e, the situation in which all employers never hire the same worker in two consecutive periods. This gives us a measure of labor fluctuation. It tells us that labor fluctuation is higher in treatments with weak EPL (31.9 % in tII and 49.3 % in tV, respectively) than in those with no EPL (16.0 % in tI and 27.0 % in tIV). Since principals cannot fire in weak EPL when re-hiring the same worker only once, this is not surprising. Of course, in treatments with strict EPL labor fluctuation is very low (6.9 % in tIII and 4.2 % in tVI).

From now on, we will count each match between an employer and a worker as one independent observation regardless of the total duration of this match, if not explicitly mentioned otherwise. This means we use, for example, the average effort spent in a match as one independent observation. Table 3.3 summarizes the number of independent matches and average match lengths, i.e., the average number of periods a match lasted, for all treatments. In the last columns of Table 3.3, we additionally distinguish between short- and long-term matches. We speak of a *short-term match*, if the match lasts for exactly one period and of a *long-term match* else. Each match is either short- or long-term. The distinction between

[13] In treatment IV, we have a fewer number of subjects (see Section 3.3). Thus, agents have to decide only 210 times. We will not always indicate similar cases in footnotes below.

short- and long-term matches is important since it divides matches in the weak EPL treatments into those with similar overall restrictions as in no EPL (short), and those where firing after the first matching period is not allowed anymore (long), i.e., that are comparable to strict EPL. It also allows us to investigate probation period effects and sanctioning behavior more thoroughly.

treatm.	numb. of obs.	Avg. mat. length	Short, freq. abs.	Short, freq. rel.	Long, freq. abs.	Long, freq. rel.
I	38	4.18	15	39.5 %	23	60.5 %
II	59	2.61	43	72.9 %	16	27.1 %
III	22	6.86	2	9.1 %	20	90.9 %
IV	47	2.96	26	55.3 %	21	44.7 %
V	85	1.85	76	89.4 %	9	10.6 %
VI	17	8.82	0	.0 %	17	100.0 %

Table 3.3: Match lengths, all treatments

Due to the high labor fluctuation, match length is lowest for treatments with weak EPL and highest for strict EPL. Sanctions increase match length for no EPL and weak EPL, but not for strict EPL.[14] In treatments with strict EPL, naturally almost all matches are long-term, with and without sanctions available. The fact that more matches are long-term in no EPL than in weak EPL suggests that principals are more reluctant to enter long-term matches in weak EPL due to the weakening of their strategic position (see also Section 3.5.2). For both of the latter degrees of EPL, the availability of sanctions increases the frequency of long-term matches, which is the intuitive result. We performed Fisher's Exact tests and χ^2-tests to check whether the differences in frequencies of long- and short-term matches between the treatments with the same degree of EPL are significant. The results show that only the difference between tII and tV is significant.[15]

As described above, workers always want to be employed and employers most

[14] This is mainly due to the fact that 6 out of 22 matches in tIII are ended by a quitting agent, while only 1 agent does this in tVI (see the following Subsection 3.5.3.2).

[15] As discussed before, it is still controversial whether to use a) Fisher's Exact tests, b) standard χ^2-tests or c) χ^2-tests with Yates-correction. In the following we give p-values for all of them in the form 'a)/b)/c)'. Treatments tI vs. tIV: $p = .191/.191/.217$; tII vs. tV: $p = .014/.014/.019$; tIII vs. tVI: $p = .495/.495/.586$.

often want to hire which largely confirms these parts of the *benchmark hypotheses*. This is not true for effort, output, sanctions, principals' earnings, and agents' payoffs. Using each match as one independent observation again, all relevant tests against the benchmark predictions, except one, are significantly different with $p < .010$.[16] In particular, effort and output are always higher than predicted in all treatments and average sanctions exceed the prediction of 0 in tI, tII, and tIII. This consequentially lets principals always earn more than predicted and agents less. The effort levels already confirm *Hypothesis 1* of our behavioral hypotheses that demands the exact opposite as the benchmark hypothesis.

Result 1: *Except for principals' hiring behavior and agents' employment behavior, the benchmark predictions are refuted.*

3.5.2 Effort levels

In each period of our experiment, groups are inevitably in different matching periods, i.e., they have just started their contract, are already matched for 3 or 4 periods, for example, or are in the last matching periods. Simply giving period averages would thus only be of limited importance, since results for different matching periods (i.e., different stages of a labor relationship) would be mixed. To avoid this problem, we distinguish between matching periods in the following instead, and thereby tolerate that the number of observations becomes lower for higher matching periods. For tI, e.g., there are 38 obs. for matching period 1 (mp1), 14 for mp5, and only 5 for mp10. This means we sometimes have only a few observations in mp10, but at least 17 in mp1, and at least 67 in the broad category subsuming matching periods 2 to 9 (mp2-9).[17] We thus mainly concentrate on mp1 and mp2-9 when discussing descriptive results below.

Table 3.4 gives the averages of efforts in total and for mp1, mp2-9, and mp10, and the averages for sanctions already.

We will first focus on effort averages on aggregate level given in the second column of Table 3.4. For both cases (sanctions available or not) efforts decrease the tighter

[16] The minor exception is in tII where principals' earnings are only slightly non-significantly higher than predicted (one sample t test $p = .051$). All other data on request.

[17] With the exception of tII (in which there are 10 observations in mp8, but only 4 in mp9 and 3 in mp10), we have at least 5 observations in mp10 in all treatments.

	Avg. effort in matching period(s)				Avg. sanction in matching period(s)			
treatm.	1-10	1	2-9	10	1-10	1	2-9	10
I	3.37	3.26	3.41	3.20	.25	.49	.18	.00
II	2.92	2.93	2.90	3.00	.38	.40	.37	.40
III	2.80	3.00	2.87	1.50	.29	.28	.30	.24
IV	3.30	3.21	3.44	2.00	–	–	–	–
V	2.02	2.51	1.48	1.00	–	–	–	–
VI	1.70	2.47	1.58	1.83	–	–	–	–

Table 3.4: Average efforts and sanctions, matching periods

the EPL becomes, which is not surprising considering the gradual weakening of the principals' strategic position. However, the decrease from weak EPL to strict EPL is rather small in both cases and the difference between average efforts in sanction treatments tI and tIII (3.37 - 2.80 = .57) is much smaller than the difference in tIV and tVI in which sanctions are not admissible (3.30 - 1.70 = 1.60). This means that the negative effect on effort levels from tighter EPL is less pronounced when sanctions as an additional disciplining option are available.

When comparing treatments with the same EPL regime, it is apparent that adding sanctions increases efforts. This effect is rather large when firing is not allowed at all (efforts in treatments with strict EPL are 2.80 and 1.70, respectively) or with weak EPL (averages are 2.92 and 2.02), but only marginal when firing is another option, i.e., in treatments with no EPL (averages are 3.37 and 3.30).

Since visual inspection gives reason to believe that the data is not normally distributed, we perform a two-way ANOVA on ranks for effort averages distinguishing between the degree of EPL on the one hand, and sanction availability on the other hand (using each match as one independent observation, again). It confirms our results summarized above: There is no interaction effect between EPL and sanctions, but both main effects are significant[18], i.e, sanctions and the degree of employment protection can be regarded as two separate factors that both significantly affect effort levels. Standard multiple comparison algorithms then tell us that, indeed, no

[18]The p-values for the two-way ANOVA on ranks are: interaction effect: $p = .366$; main effect of EPL: $p = .002$; main effect of sanctions: $p = .004$

EPL as a separate factor differs from weak EPL and strict EPL, but weak EPL and strict EPL do not.[19] Since the two-way ANOVA on ranks is sometimes criticized (see, e.g., Toothaker and Newman (1994)), we additionally perform two one-way non-parametric equivalents of the ANOVA, this means two Kruskal-Wallis tests with the multiple comparison algorithm introduced by Steel, Dwass, Critchlow, and Fligner (see, e.g., Critchlow and Fligner (1991)). They tell us that efforts are not statistically different in tI-III, but do differ in tIV-VI in the same way as before: efforts in tIV are higher than those in tV and tVI.[20]

Finally, we perform one-sided two sample tests between the treatments with the same degree of EPL.[21] We use Welch-Satterthwaite t tests (*WS tests*) for samples large enough, and Mann-Whitney U-tests and robust rank-order tests *(U-tests (rro tests))* for treatments tIII and tVI in which visual inspection indicates that distributions are skewed differently. The test results confirm the descriptive impressions: efforts in tI and tIV do not differ significantly, while those in tII and tV as well as in tIII and tVI, respectively, do.[22]

In summary, all these results strongly confirm *Hypotheses 2a and 2b*.

Result 2: *With and without sanctions, efforts decrease the tighter the EPL is. This trend is stronger when sanctions are not available. Sanctions largely increase effort when comparing the two treatments with weak EPL or the two treatments with strict EPL. There is almost no increasing effect of sanctions when firings are allowed (no EPL-treatments).*

Figure 3.1 further details the results for all six treatments. In this figure, treatments are visually distinguishable by two factors: the sanction availability is indicated by lines (solid lines for tI-III, dashed lines for tIV-tVI), while the degree of EPL is depicted by filled and unfilled symbols (diamonds for tI and tIV with no EPL, triangles for tII and tV with weak EPL and circles for tIII and tVI).

[19] Results for Bonferroni (Tamhane) are: no EPL vs. weak EPL: $p = .011(.008)$; no EPL vs. strict EPL: $p = .006(.001)$; weak EPL vs. strict EPL: $p = .807(.514)$

[20] Kruskal-Wallis for tI-tIII: $p = .240$ – Kruskal Wallis for tIV-VI: $p = .002$; multiple comparisons with Steel, Dwass, Critchlow, Fligner: tIV vs. tV: $p \approx .025$; tIV vs. tVI: $p < .001$; tV vs. tVI: $p > .200$.

[21] The respective hypothesis is also explicitly one-sided

[22] WS tests : tI vs. tIV: $p = .233$; tII vs. tV, $p = .014$; U-test (rro test): tIII vs. tVI, $p = .006(< .001)$.

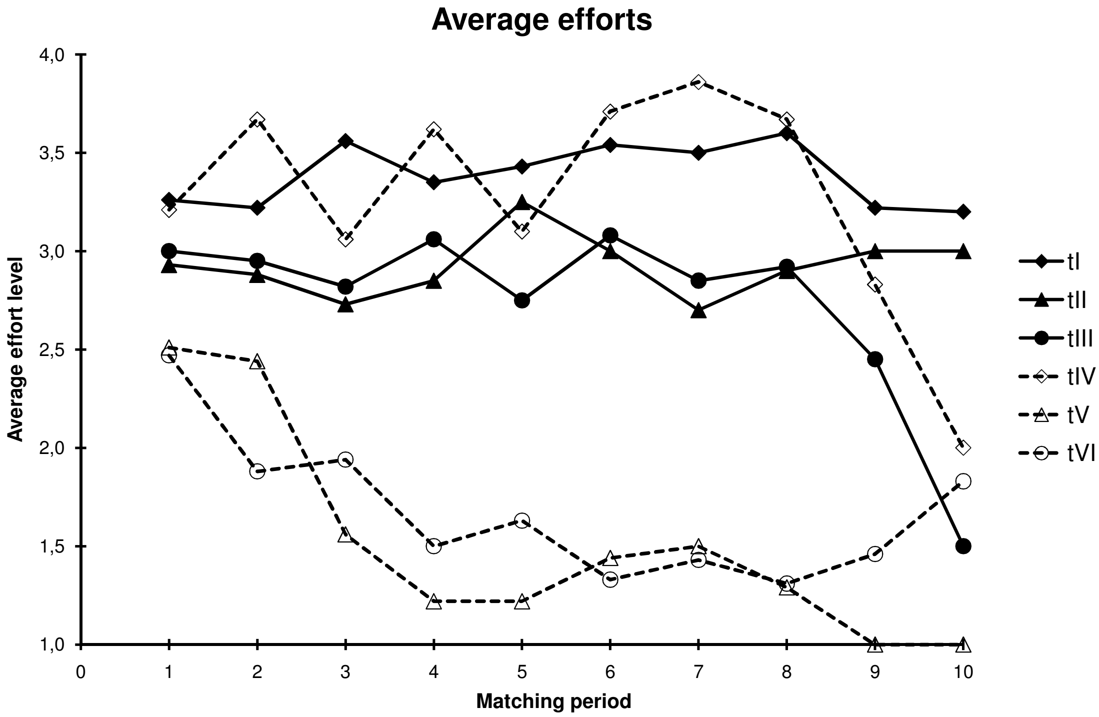

Figure 3.1: Effort choices, matching periods

First of all, notice that Figure 3.1 again confirms that efforts are larger than the minimum, except for mp9 and mp10 in tV. In the first matching period, a) efforts are similar in tI and tIV (3.26 and 3.21), b) a little higher in those two than in tII and tIII, which are also quite similar (2.93 and 3.00), and c) somewhat lower in tV and tVI than in all other treatments (2.51 and 2.47). This ordering stays the same during mp2-9, but while efforts hardly change in tI-IV, they quite sharply decrease in tV and tVI (see also Table 3.4). Endgame effects in mp10 only occur for tII, tIV and tV, which might be of less importance due to the small number of observations.[23] Overall, when mainly focusing on mp1-9, these findings confirm the results discussed above and, additionally, give reason to believe that the expected probation period effect as formulated in *Hypothesis 2c* exists for tV. We will elaborate on this in the following paragraphs. They also indicate a surprisingly

[23]In fact, when analyzing periods instead, large endgame effects are clearly visible for all treatments, except tVI where efforts are lowest already. Figures on request.

sharp decrease in effort levels after mp1 for tVI, which kind of puzzles us. Maybe participants feel a little honored by being (randomly) chosen to work, while others have to be unemployed and thus spend more effort in mp1, but quickly forget this gratitude when realizing their strong strategic position.

Let us now distinguish between short- and long-term matches. In Figure 3.2 long-term average efforts are represented by lines as before, but only for mp1, mp10 and the broad category of mp2-9, since this suffices here. We additionally give the averages in short-term matches, i.e., in matches that ended after one matching period, by short horizontal lines in this figure, but restrict this information to tI, tII, tIV, and tV, since in the two treatments with strict EPL almost all (tIII) or all matches (tVI) are long-term. The specific short-term averages are $e = 3.00$ for tI, $e = 2.86$ for tII, $e = 2.58$ for tIV, and $e = 2.42$ for tV.

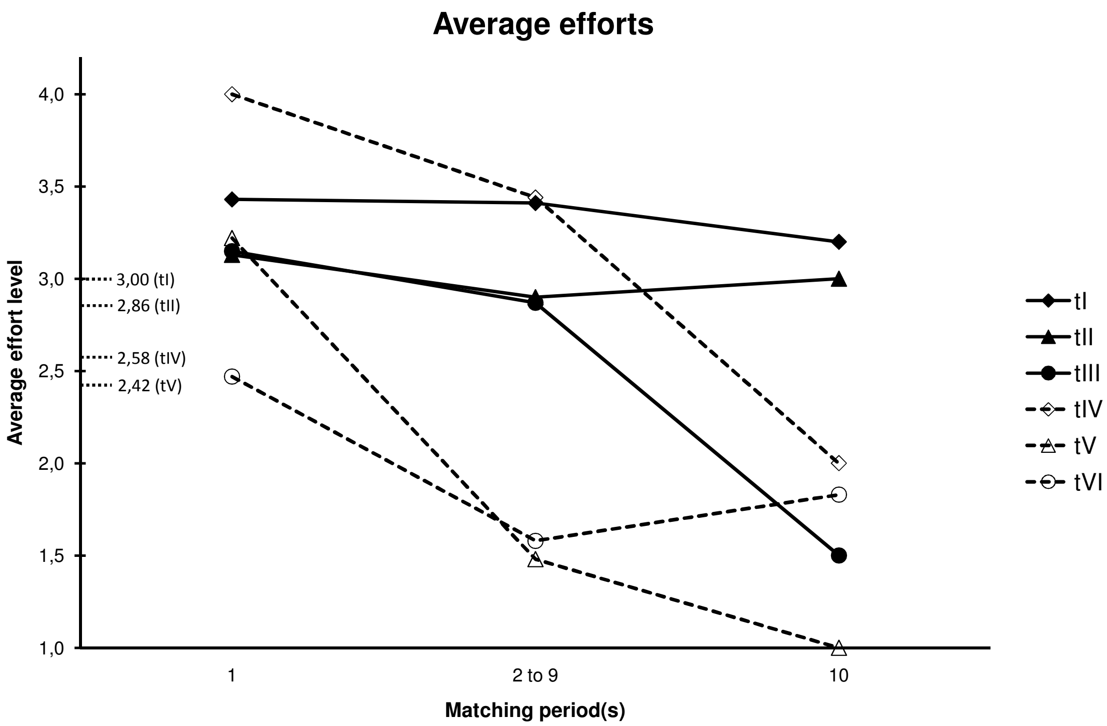

Figure 3.2: Effort choices, long- and short-term matches

When disregarding mp10 due to the few number of cases, there are mainly two insights delivered by Figure 3.2. Firstly, for the sanctions treatments tI-III, average efforts in long-term matches do not change that much and are (a little) higher than those in their short-term counterparts, but, there are quite sharp decreases in average efforts in the treatments tIV-VI in which sanctions are not available. Secondly, as expected, the decrease is highest for tV (from about 3.2 to about 1.4), where the long-term average effort in mp1, indeed, lies above the short-term average (of about 2.4), but this order changes in mp2-9, although now we are comparing different matching periods. On the contrary, efforts in tII decrease only a little from about 3.15 to about 2.90. We perform dependent sample tests comparing the averages in mp1 to those in mp2-x, where x stands for the last matching period of the long-term match, for all treatments, mainly to test whether the *probation period effect* for tV just reported is significant. Indeed, all (modest) decreases are significant, except for the one in tII.[24] This partly confirms *Hypothesis 2c*. We conclude:

Result 3: *Efforts decrease in long-term matches after the first matching period, but with sanctions available the magnitudes of these effects are negligible. There is a strong probation period effect in tV.*

3.5.3 Sanctions

Table 3.5 details sanction frequencies, distinguishing between those matches in which principals sanction at least once and those in which they do not.

	How often do principals sanction?				
	Never		At least once		
treatm.	abs. freq.	rel. freq.	abs. freq.	rel. frq.	sum
I	9	23.7 %	29	76.3 %	38
II	21	35.6 %	38	64.4 %	59
III	2	9.1 %	20	90.9 %	22

Table 3.5: Sanction frequencies

In all three sanction treatments, we observe many more principals who sanction

[24]We generally performed two-sided Wilcoxon paired sample tests, since we did not expect probation period effects for all treatments. Results are: tI: $p = .029$; tII: $p = .196$; tIII: $p = .048$; tIV: $p < .001$; tV: $p = .016$; tVI: $p = .005$.

at least once than principals who never sanction. Most never-sanctioning principals can be found in tII (35.6 %), followed by tI (23.7 %) and tIII (9.1 %). Obviously, sanctions are used by more principals than the zero prediction of the benchmark model, even though they are costly. Consequentially, binomial tests refute the hypothesis that no principal sanctions at least once for all three treatments with $p < .001$. The first part of *Hypothesis 3* is thus confirmed.

Result 4: *Sanctions are frequently used by principals in tI, tII and tIII.*

Furthermore, Fisher's Exact tests and χ^2-tests comparing each of the two treatments with sanctions with regard to their frequency of never-sanctioning principals, tell us that the differences between tI and tII as well as tI and tIII, respectively, are not significant, but those for tII and tIII are.[25] This means that the frequency of principals who sanction does not continually increase the tighter the EPL becomes, which partly rejects *Hypothesis 4*.

Figure 3.3 and the previously-introduced Table 3.4 give sanction averages for treatments I-III, distinguishing between mp1, mp2-9 and mp10. In this figure, lines are differently dashed than before for reasons of visual differentiation.

The most immediate result suggested by Table 3.4 and Figure 3.3 is that on aggregate level, i.e., for all matching periods, average sanctions are indeed larger in tII with weak EPL ($s = .38$) than in tI with less tight EPL ($s = .25$), but this trend does not continue to strict EPL. In particular, average sanctions there decrease again to $s = .29$. Overall, sanction differences are not very large and the Kruskal-Wallis test for differences in sanction averages in tI-III is insignificant with $p = .915$. These aggregate results do not suggest that sanctions consistently increase the tighter the EPL becomes.

However, distinguishing between matching periods reveals several interesting details. The sanction ordering in the first matching period follows the exact opposite pattern than expected: Average sanctions in mp1 are highest in tI ($s = .49$) and (a little) lower in tII ($s = .40$) and tIII ($s = .28$). Differences are not significant again.[26] The order changes in later matching periods, almost exclusively because sanctions sharply decrease in tI to $s = .18$ in mp2-9 and $s = .00$ in mp10. Changes

[25] The p-values are given in the same form as in footnote 15. Treatments tI vs. tII: $p = .264/.264/.311$; tI vs. tIII: $p = .189/.189/.288$; tII vs. tIII: $p = .025/.025/.038$.

[26] The p-value of the Kruskal-Wallis test restricted to mp1 is $p = .549$.

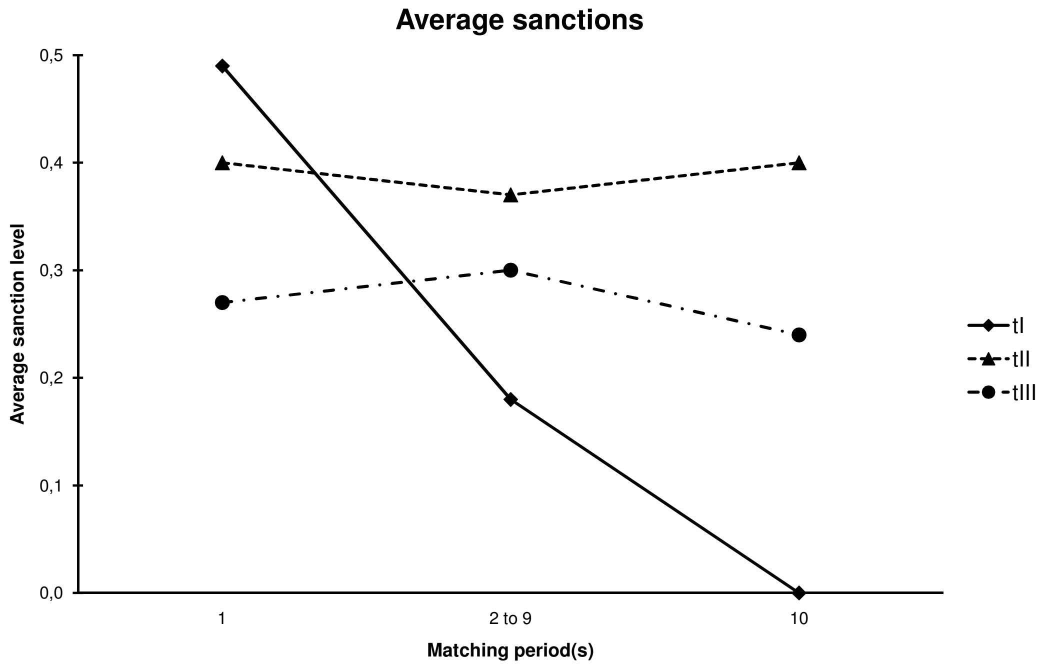

Figure 3.3: Sanctions, matching periods

are negligible for tII and tIII, which means that sanction averages in mp2-10 are highest in tII, followed by tIII and tI. Since we have only 3 observations in mp10 of tII, we limit ourselves to discuss mp1-9 from now on.

Let us take a deeper look at the crucial role the weak EPL treatment tII plays in understanding what happens during our experiment. In the first matching period, no EPL and weak EPL factually induce the same degree of employment protection. By contrast, in mp2-9 the EPL rules in weak EPL are the same as those in strict EPL. We denote the sanction s in matching period(s) mp of treatment t with $s_{t,mp}$ now. When detailing our aggregate hypothesis that average sanctions increase the tighter the EPL becomes, to single matching periods, it thus seems intuitive to expect (O1) $s_{tIII,m1} > s_{tII,m1} \approx s_{tI,m1}$ to hold in mp1, while in mp2-9 we should expect (O2) $s_{tIII,m2-9} \approx s_{tII,m2-9} > s_{tI,m2-9}$. In fact, our results for (O2) at least follow the expected order, since we observe similar sanctions for tII and tIII ($s_{tII,m2-9} = .37$ and $s_{tIII,m2-9} = .30$) that are a little higher than those in tI ($s_{tI,m2-9} = .18$). However, (O1) is clearly violated since, indeed, average sanctions are quite similar in tI and tII ($s_{tI,m1} = .49$ and $s_{tII,m1} = .40$) as expected, but instead of being lower than those in strict EPL, they are higher ($s_{tIII,m1} = .28$). We will now present further analyses investigating this puzzle, but for the time being let us shortly summarize the refusal of *Hypothesis 4*.

Result 5: *Differences in average sanctions are rather small. Average sanctions are largest in tII with weak EPL, followed by tIII with strict EPL and tI with no EPL. Sanctions are frequently used in all treatments. Sanction frequencies are largest in tIII, but do hardly differ between tI and tII.*

Distinguishing between averages for short- and long-term matches gives an important hint why sanctions in mp1 might be higher for tI and tII than expected. Figure 3.4 is constructed analog to the previous Figure 3.2 that detailed average efforts in short- and long-term matches. Short-term averages are again given after short horizontal lines and tIII is left out, since there almost all matches are long-term. The short-term averages are $s = .69$ for tI and $s = .51$ for tII.

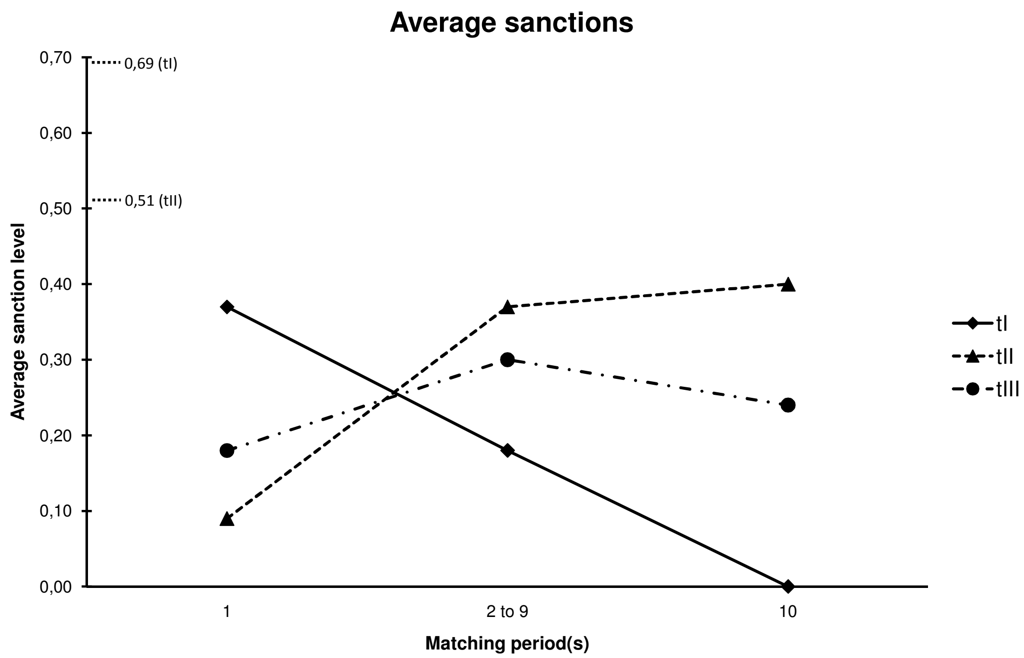

Figure 3.4: Sanctions, long- and short-term matches

Long-term match averages are almost the same as on aggregate level (i.e., for the sum of all matches, short or long-term), with the exception that in mp1 sanctions in tII are quite low. Wilcoxon paired sample tests comparing the sanction average in mp1 with that of mp2-x tell us that only the increase in average sanctions in tII is significant, which probably is the reason for the relatively mild probation period effect in tII discussed above.[27]

Much more illuminative are the differences between long-term and short-term sanctions in mp1: Here, sanctions in long-term matches of tI are only about half as high as in their short-term counterparts ($s = .37$ and $s = .69$), while this trend is even more pronounced for tII ($s = .09$ and $s = .51$). Mann-Whitney U-tests tell us that the difference in tI is insignificant on the 10%-level ($p = .135$), but significant for tII ($p = .006$). By definition short-term averages are those of matches that end after the first period. As will be shown below, most of these endings are caused by the principal firing his employer. Knowing this, our interpretation of the results given by Wasmer (2006) that firings and sanctions serve as substitutes is refuted by our findings. Instead, it seems as if the principals who fire their agents (short-term matches) additionally sanction them, i.e., there is reason to believe that sanctions and firings are used as complements.

A loosely comparable result has been discussed in experimental literature: In a setting in which costly sanctions and indirect reciprocity could have been used concurrently, Rockenbach and Milinski (2006), to their own surprise, found them both to be used simultaneously.

Now, we further investigate the relation between firings and sanctions in our experiment by moving our attention to firing behavior.

3.5.3.1 Firing behavior

Note that in tIII and tVI there cannot be firings by definition, in p10 of all treatments, firings are meaningless and thus not allowed by our software, and in tII and tV, principals are only allowed to fire during the first matching period. Thus, the following Table 3.6 gives matching period averages only for observations from periods 1-9 in which firings are de facto allowed, divided into those without firing and with firing. This Table treats each matching period as one observation (henceforth, *'mp-rule'*) which has the disadvantage that these observations are not independent

[27]The p-values are: tI: $p = .852$; tII: $p = .002$; tIII: $p = .092$.

from each other. This means that we can give descriptive statistics, but cannot always perform tests. This leaves the question why we do not stick to the 'one match = one observation'-rule (*'one match-rule'*) we used before that does not have this disadvantage.

Let us answer this with an example: Suppose a principal in tI fires her worker in matching period 7 of a match and thus decided against firing in mp1-6. When using the 'one match-rule', we would nevertheless count all these 7 single observations – that are likely to differ a lot in effort decisions etc. – as one observation with firing. Instead, we categorize each matching period either as one with firing or without, implicitly assuming that this last matching period in which the principal fires is the principal's most important stimulus for the firing decision. If one instead thinks that the average of efforts, outputs etc. is relevant for the firing decision, one should use the 'one match-rule'. We will also give results for the 'one match-rule' below. They do not qualitatively differ from those with the 'mp-rule'.

	treatm. I		treatm. II		treatm. IV		treatm. V	
P fires?	no	yes	no	yes	no	yes	no	yes
# of obs.	129	14	16	40	92	34	9	69
Effort	3.50	3.07	3.12	2.85	3.70	2.91	3.22	2.55
Output	3.71	2.79	4.06	2.95	4.27	2.82	3.11	3.36
Sanctions	.24	.61	.09	.48	–	–	–	–
Earnings (P)	18.32	13.31	20.21	14.28	21.36	14.12	15.56	16.81
Payoffs (A)	10.03	9.25	10.67	9.69	10.67	11.15	10.92	11.17
Outp. dev.	.05	-.42	.48	-.35	.26	-.70	.06	.20

Table 3.6: Efforts, Outputs, Sanctions, and Earnings, with and without firing

The high labor fluctuation in treatments with weak EPL described above immediately follows from the number of observations given in the third row of Table 3.6. While in tI and tIV principals most often hire the same agent in consecutive periods (tI: 129 out of 143 times, i.e., in 90.2 % of cases, treatIV: 73.0 %), principals in tII and tV most often fire (frequencies, calculated as before, are: tII: 28.6 %, tV: 11.5 %). As expected, efforts are lower in matching periods in which the principals eventually fired their agents.[28]

[28]Except for sanctions the other results of this table are discussed in later sections.

Averages for sanctions confirm our intermediate findings that sanctions are rather used as complements than as substitutes. In periods of tI in which the principal fires his agent, average sanctions are $s = .61$, compared to only $s = .24$ where she hires him again. For tII the difference is even more striking: $s = .48$ for firing and $s = .09$ for re-hiring principals. Overall, principals sanction agents more severely in the periods they also fire them. As discussed above using the 'mp-rule' not always allows for statistical testing, since for tI, for example, the sanction averages in matching periods in which the principals fire, are not independent from those in which they do not fire. For weak EPL, this problem does not occur; here, principals have to fire in the first matching period and we can compare averages in this first matching period with the averages in the first matching period of matches in which principals do not fire. The difference in average sanctions is highly significant between firing and non-firing matching periods of tII (U-test: $p = .008$).

The sanction averages just discussed give reason to believe that sanction severeness is lower in matching periods in which the principal fires when using the 'mp-rule'. From now on, we use the 'one match-rule' again, if not explicitly mentioned otherwise. Table 3.7 categorize each match in either one that ends with the principal firing the agent or not. It also further distinguishes the fractions of principals who never sanction and those sanctioning at least once initially given in Table 3.5 to matches with and without firing.[29]

		How often do principals sanction?				
		Never		At least once		
treatm./fired?		abs. freq.	rel. freq.	abs. freq.	rel. frq.	sum
I	yes	5	35.7 %	9	64.3 %	14
I	no	3	13.6 %	19	86.4 %	22
II	yes	19	47.5 %	21	52.5 %	40
II	no	3	18.8 %	13	81.2 %	16

Table 3.7: Sanction frequency, with and without firing

In most of the matches without firing, principals choose to sanction at least once

[29]We exclude the two (tI) respectively three matches (tII) that start in period 10, since these are the only ones in which the principals could never fire.

(86.4 % in tI and 81.2 % in tII), but the majority also does this in matches that end with a firing, although the relative frequencies are a little lower this time (64.3 % in tI and 52.5 % in tII). The high percentages of principals that sanction and fire, again, suggest that sanctions and firings are both used simultaneously. With Table 3.7 alone, one might argue that maybe principals in tI first used sanctions as a disciplinary device and then, in the last matching period, completely substituted them by firings, but the data presented in Table 3.6 doubted this possibility. Further confirmation comes from looking at the descriptive data of the long-term matches in tI that end with firings. In those 6 long-term matches the sanction average in all matching periods except for the last one is $s = .49$, while it even increases in the last matching period to $s = .75$, i.e., principals in long-term matches sanctioned their agents even harder in the last matching period in which they also fired them. This clearly does not indicate the substitutive use of sanctions and firings, even though the difference is insignificant.[30]

Similar conclusions follow for sanction averages on match level within the same treatment: In tI (tII) the average sanction of the 22 (16) matches without firing is $s = .48$ ($s = .39$), but even higher instead of lower with firing: $s = .53$ ($s = .48$).

Finally, Table 3.8 allows us to discuss yet another aspect of the relationship of firings and sanctions. Here, it is distinguished between four average sanction categories (no sanctions (zero), small, medium and high sanctions) and firings.

treatm.	fired?	Average sanctions, s: zero $s = 0$	small $0 < s < .5$	medium $.5 \leq s < 1$	high $1 \leq s \leq 2$	sum
I	yes	5 (35.7%)	1 (7.1%)	3 (21.4%)	5 (35.7%)	14
I	no	2 (9.1%)	12 (54.5%)	4 (18.2%)	4 (18.2%)	22
II	yes	19 (47.5%)	6 (15.0%)	4 (10.0%)	11 (27.5%)	40
II	no	1 (6.3%)	8 (50.0%)	5 (31.3%)	2 (12.5%)	16

Table 3.8: Sanction severeness categories, with and without firing

[30]Two-sided Wilcoxon paired sample test: $p = .875$.

While for firing principals, sanctions are rather bi-modal in the extreme categories ($s = 0$ and $1 \leq s \leq 2$), i.e., principals either sanction hard or not at all, they roughly resemble right-skewed normal distributions for non-firing principals. Here, the large majorities sanction moderately. There is some evidence that frequencies differ significantly.[31] This may be seen as a hint that some principals use firings and sanctions as substitutes, but others as complements.

Possible reasons for such behavior can, of course, only be speculative. In a neuroeconomical study de Quervain et al. (2004), analyzed whether activations in the dorsal striatum are an indicator of sanctioning behavior in a trust game. Sanctioning other participants incurred factual costs for the sanctioner and thus likely negative utility. Nevertheless, de Quervain et al. (2004) found that participants used sanctions. The authors also reported that participants with stronger activations sanctioned harder. They argued that the activation levels reflected anticipated positive utility from sanctioning that might outweigh the costs of sanctioning. Transferred to our case this might be interpreted in a way that some of our participants are satisfied with the costless firing while others needed the additional costly sanctioning.

Result 6: *There is evidence that at least a large fraction of principals uses sanctions as a complement to firings, rather than as substitutes.*

3.5.3.2 Quitting agents

Due to the labor supply excess and the fact that at first principals decide whether to fire or not, we do not expect that as many matches are ended by workers who quit as by employers who fire – except for treatment tIII in which firings are not allowed. Hereafter, we use *"to quit"* and *"to resign"* synonymously.

Since agents cannot quit in the last period of the experiment, the following figures and tables exclude these data.[32] Table 3.9 gives absolute (and relative) frequencies of matches that end with firings by principals, and matches that end after the agent

[31] According to Cramér's V (or standard contingency coefficients) sanction frequencies for firing and non-firing principals differ in both treatments: tI: $p = .022(.022)$; tII: $p = .001(.001)$. However, expected cell frequencies are smaller than 5 in 3 out of 8 cells, which is often cited as an heuristic for the applicability of the above-mentioned tests.

[32] To resign from the job would be meaningless here, because the experiment ends after p10.

quits. In the last two rows, the sanctions averages for matches the agent ended by quitting and all other matches are given. Consequentially, we limit ourselves here to the sanction treatments.

	Treatment		
	I	II	III
tot. numb. o. matches	36	56	22
tot. numb. o. firings (f)	14 (38.9%)	40 (71.4%)	0 (.0%)
tot. numb. o. quitting agents (q)	8 (22.2%)	3 (5.4%)	6 (27.3%)
avg. sanc., not ended by q	.34	.44	.27
avg. sanc., ended by q	1.04	.58	.74

Table 3.9: Frequencies of firing employers and quitting agents

As expected, agents do not quit as regularly as they are fired in tI (38.9 % matches end with a firing, but only 22.2 % with a resignation) and especially in tII (71.4 % and 5.4 %), but quite often in tIII (27.3 % resignations). Following the arguments introduced by Wasmer (2006), we should observe the largest differences in average sanctions between matches in which the agent does not quit eventually, and matches in which he quits in tIII – or at least a much larger fraction of voluntary resigning agents in tIII than in all other treatments. In fact, the differences are rather large in tIII: agents who do not quit are only sanctioned with $s = .27$ on average, while the others are sanctioned (bullied) more severely with $s = .74$. This difference is significant despite the small sample sizes (U-test: $p = .046$). However, the difference in average sanctions between matches in which the agents quit and those in which they do not, is even larger for tI ($s = .34$ and $s = 1.04$, respectively) and again significant (same test: $p = .009$). We additionally computed Cramér's V for the 2x2-table that distinguishes between tI and tIII on the one hand, and matches that end by an agent's resignation and those without resignation on the other hand. The differences are not significant according to this test ($p = .756$). This refutes the hypothesis that sanctions are used differently in treatments tI and tIII (without and with firing restrictions). Probably due to the small sample size, the sanction differences in tII ($s = .44$ and $s = .58$) are not significant.

Finally, we want to give the averages when again using each matching period as one observation ('mp-rule'), since the reason why these descriptive results are interesting is intuitively even more appealing this time: the most important sanctions from the agent's point of view are likely those in the matching period he quits. For tI we have got 121 matching periods in which the agent does not resign and an average sanction of $s = .19$ and 8 resignations with an average sanction of $s = 1.09$. For tII and tIII, respectively, we observe $s = .29$ (95 obs.) and $s = 1.17$ (3 obs.) as well as $s = .26$ (129 obs.) and $s = 1.17$ (6 obs.). Overall, these differences are larger than those on match level, but tell a similar story: in tI (, tII) and tIII sanctions are triggering voluntary resignations.

Result 7: *Agents quit regularly in tI and tIII and not very often in tII. Agents who quit are sanctioned more severely than those who do not quit.*

3.5.3.3 Output, output deviations and correlations

We introduced uncertainty on behalf of the principal concerning the effort choices. We did this because we think it makes EPL meaningful. Under perfect monitoring, there would be little room for opportunistic behavior: for example, contracts could be contingent on work effort and underachieving workers would probably not be protected by employment protection laws.

In our model, higher effort induced higher expected output. In Table 3.6, using the 'mp-rule', we already gave output averages and output deviations, defined as the difference between own and average output, for the treatments with firing opportunity using each matching period as one observation. It seems likely that both variables are most important for the principal in the immediate period she fired. The output averages mirror the effort choices in tI, II and IV, but are a little higher in tV for the principals who do fire. This is probably due to the small sample size of non-firing principals. The same holds for output deviations. They are positive for non-firing principals, but negative for non-firing ones, except for treatment tV in which they are about equal.

We now want to check the correlations between these results just presented for outputs and the sanctioning by principals using the 'one match-rule'. We predicted the sanctions to be more severe for smaller output deviations (*Hypothesis 3*). In fact, for tI and tII Spearman's rank correlation coefficients for the relationship of sanctions and output deviation are negative ($\rho = -.611$ and $\rho = -.421$) and highly significant ($p < .001$ and $p = .001$). This means that principals sanction the harder the lower the own output in comparison to the average output is.

Similar results hold when using output or principals' earnings instead.[33] For tIII, however, the correlation between output deviation and sanctions still has the same sign, but is much lower ($\rho = -.021$) and becomes insignificant ($p = .926$). Again this tendency holds for output and earnings.[34]

Result 8: *Sanctions and output deviations are negatively correlated in tI and tII, but not in tIII.*

3.5.4 Payoffs

For the sake of completeness, let us finally just summarize the main results for principals' earnings, π, and agents' payoffs, ω. We will mainly limit ourselves to the descriptive data for mp1-9 given in Table 3.10, and will just highlight the important findings for the sub-groups of firing and non-firing principals (see Table 3.6).

	Avg. earnings Principals matching period(s)			Avg. payoff Agents matching period(s)		
treatm.	1	2-9	10	1	2-9	10
I	16.61	18.78	22.00	9.46	10.25	10.80
II	15.87	17.29	12.93	9.87	10.04	9.97
III	17.90	17.43	19.26	10.24	10.26	11.11
IV	17.13	20.35	18.33	10.94	10.82	11.46
V	16.41	12.69	12.00	11.21	11.79	12.00
VI	16.47	15.95	15.42	11.31	11.77	11.63

Table 3.10: Average payoffs, matching periods

[33] Output: tI (tIV): $\rho = -.795(-.497)$, $p < .001(< .001)$; Principals' earnings: tI (tIV): $\rho = -.844(-.656)$, $p < .001(< .001)$

[34] Output: $\rho = -.185$, $p = .410$; Earnings: $\rho = -.284$, $p = .200$

Columns 2-4 of Table 3.10 detail the average of principals' earnings in the first matching period, in mp2-9, and mp10. While principals earned about the same in mp1 in all six treatments, their earnings increase in the treatments with no EPL (tI and tIV), stay about the same in the remaining treatments with sanctions (tII and tIII) and (moderately) decrease else in mp2-9. As could be expected, earnings are highest in the four treatments with firing and/or sanction opportunity now. Since sanctions are used and costly, principals earn most in tIV with $\pi = 20.35$, followed by the only treatment with allowed firing (namely tI with $\pi = 18.78$), and the two other sanction treatments tII ($\pi = 17.29$) and tIII ($\pi = 17.43$). Earnings are a little lower in tVI ($\pi = 15.95$) and the large decrease in tV to $\pi = 12.69$ is due to the probation period effect that cannot be offset by sanctions here.

The distinction between firing and non-firing principals done in Table 3.6, using the 'mp-rule', shows us that firing principals in tI, tII, and tIV earn much less than those that do not fire, while this is not the case in tV (probably since principals most often fired here). For tII we can test this difference directly when restricting attention to mp1, and it is significant.[35] We again choose the indirect way with dependent sample tests for long-term matches for tI and tIV using the 'one-match rule'. Due to the small sample size, the large difference in tI ($\pi = 17.12$ in the periods before firing and only $\pi = 9.25$ in the firing period) is slightly insignificant unlike the even larger difference in tIV ($\pi = 22.82$ and $\pi = 12.78$).[36]

For agent payoffs, it suffices to summarize that a) on average, they are about the same in the sanction treatments, a little higher in tIV and, of course, highest in tV and VI (see Table 3.10); b) fired agents in tI and tII earn less than those not fired (see Table 3.6) which, again, indicates that they are fired and sanctioned at the same time (especially since fired agents save costs by spending less effort); and c) this is not the case for fired agents in tIV and tV who, in fact, earn more in the firing period than the non-fired ones.

[35] U-test (rro test): $p = .001 (< .001)$

[36] Two-sided Wilcoxon paired sample tests: tI (6 obs.): $p = .063$; tIV (9 obs.): $p = .004$.

3.6 Conclusion

The employment effects of employment protection legislation are the main focus of most scientific works on this topic (see, e.g, Hopenhayn and Rogerson (1993), Mortensen and Pissarides (1994) or Garibaldi (1998)). Wasmer (2006) shifted attention to another interesting aspect of employment protection: the question whether firms that are not allowed to fire due to a stricter degree of EPL react by replacing the missing firing incentives by bullying their workers. In a theoretical model and in an empirical analysis, using, amongst other things, the number of prescribed anti-depressants as a proxy for bullying at the workplace, Wasmer (2006) found such a relation between firings and bullying.

In our study, we tried to establish a more direct test of the interplay between employment protection legislation and bullying by designing a laboratory experiment in which employers could costly sanction their employees. We interpreted the costly sanctions as bullying and thereby connected the work of Wasmer (2006) with the experimental study on the relation between employment protection and rewards by Falk, Huffman and MacLeod (2008), and the existing literature on sanctions and rewards (see, e.g., Andreoni, Harbaugh and Vesterlund (2003)).

We found that 1) the possibility to sanction, ceteris paribus, increased effort levels when comparing treatments with the same degree of EPL and 2) effort averages were larger in treatments in which firings were not prohibited. This could have been expected, since it is in accordance with the literature on sanctions and the results of Falk, Huffman and MacLeod (2008).

Surprisingly, we also found that 3) sanctions did not continually increase the tighter the employment protection became and that, on the contrary, 4) average sanctions were almost equally high in both extreme treatments: the treatments in which firings were always allowed and always prohibited, respectively.

We presented evidence suggesting that this was probably due to the fact that sanctions and firings were used as complements, rather than as substitutes in our experiment – at least by a considerable fraction of principals. This result refuted the hypothesis that bullying was primarily used to replace firing incentives, but loosely resembled an insight discussed by Rockenbach and Milinski (2006).

4 Downsizing the labor force

4.1 Introduction

Downsizing is often understood as laying off a large group of workers and has become a synonym for, allegedly unfair, firing decisions in popular (science) literature. More broadly, downsizing could refer to all situations where a subgroup of interacting parties within an organization would have better success or survival prospects than the whole group. Dramatic examples of lifeboats with too little water or food for all can be found in novels and are hopefully more often fictional than factual. One reaction to such challenges could be volunteers offering to be excluded. But such hero volunteers are probably a rare species. What one realistically has to expect are attempts of some parties to exclude others against their will. This does not only invoke material aspects but also raises moral and emotional concerns not only of those who suffer, e.g., by being excluded, but also of those who exclude others.

One can hope to capture crucial aspects of firms' downsizing decisions by employing rather abstract scenarios like ultimatum games, e.g., with one proposer and several responders of whom some can be excluded (Fischer et al. (2008)). But then questions like *"Why can the proposer and some responders exclude other responders?"* or *"Why can the remaining players share more, and how is that related to what all would receive?"* naturally arise. These are less troublesome when considering situations of which downsizing is typical, namely a firm which tries to reduce its labor force, although it is prospering.

This suggests a principal-agent setting and that the initiative for downsizing naturally rests on the owners, respectively their delegates, e.g., CEOs. What can be gained by downsizing is then implied by the economic, technological, and legal environment. A smaller labor force may avoid bankruptcy, resembling the example of a lifeboat whose supplies do not suffice for all. Investigating such situations in which either all suffer or some survive may yield important insights. But here we focus on situations in which downsizing is not a matter of immediate necessity but one of profitability. Such downsizing announcements regularly alert the public all over the world (New York Times (2008)), e.g., in Germany in the recent past. In January 2008, the mobile phone producer Nokia announced to shut down a factory in Bochum, Germany, and lay off 4,300 full-time employees and temporary workers, although internal accounting showed a profit of € 134 million.[1] In February 2008, the household products company Henkel and the automobile manufacturer BMW announced layoffs of 3,000 and 8,100 workers, respectively, although their profits had increased (to about € 1 billion and more than € 3.75 billion, respectively). These firms justified downsizing by future risks due to the Global Economy (Henkel) or simply by higher rentability aspirations (BMW).[2]

We want to study downsizing experimentally, although the external validity of such experiments for life-altering payoffs might be criticized. We admit that real layoffs might induce dramatic consequences that cannot be perfectly reproduced in the lab, but are confident that behavioral aspects of layoffs can at least be revealed in its main tendencies. More specifically, we experimentally examine the occurrence and the behavioral effects of downsizing in a principal-agent setting where workers chose effort levels. Our setting thus fundamentally differs from the ultimatum game environment Fischer et al. (2008) use. We pick up the labor market institution of minimum wages again that serves as a labor market rigidity to allow for profitable downsizing. In a treatment in which a principal's profit gain from downsizing his labor force is positive, but rather small, we expect layoffs less often than in another treatment in which this gain is rather large.

Other research questions we are going to answer are: Do game theoretic benchmark solutions predict contract offers and agents' behavior?, Are there probation period effects like in the employment protection experiment in Chapter 3?, and,

[1] Nokia executives stated that these internal numbers are due to accounting regularities and do not represent the factual profitability of the factory at all (see FAZ (2008b)).

[2] See, e.g., FAZ (2008a), Handelsblatt (2008) or Frankfurter Rundschau (2008) for detailed information concerning the downsizing announcements, profits, and reactions.

Will there be differences between anticipated and unanticipated downsizing and, if so, will they question the predicted profitability of downsizing?

The major finding we will report is that downsizing does not occur more regularly in the treatments with larger theoretical and factual gains from firing than in the other treatments. We are going to report some findings suggesting that this might be caused by the fact that workers (intuitively) spend comparatively more effort when firing incentives are high, especially when they know that employers are allowed to fire them later on and during probation periods. Within a specific treatment, the employer participants who downsize their labor force are those with the lowest profits. Benchmark predictions deduced from a model with pure payoff-maximizing players are shown to be rather useless. In particular, piece rate offers are surprisingly low.

The specific firm model, which we analyze theoretically and have implemented experimentally, is introduced in Section 4.2 together with its solution. Section 4.3 describes the experimental protocol and Section 4.4 discusses the hypotheses. After analyzing the data in Section 4.5, the conclusion follows in Section 4.6.

4.2 The principal-agent model

■ **Model description:** Principal P currently employs both, a highly productive agent (agent h) with cost function, $C_h(e_h)$, of effort e_h and a less productive one (l) with cost function $C_l(e_l)$. Both produce the same kind of output. The highly productive agent has lower costs of effort, i.e., $C_h(e)$ is smaller than $C_l(e)$ for all positive effort levels e. More specifically, we rely on quadratic effort cost functions $C_h(e_h) = \frac{k}{2}e_h^2$ and $C_l(e_l) = \frac{d}{2}e_l^2$ with $0 < k < d$. Each unit of effort results in one additional unit of output with the principal perfectly observing the agents' types and the amount of output produced.[3]

To allow for profit-increasing downsizing, we assume the labor market rigidity

[3]One can justify co-employment of more and less productive workers by new production techniques which are more easily adopted by some, e.g., the younger workers, but this questions the productivity of others who before were equally skilled. Many other principal-agent models use the same or similar convex functions (see, e.g., Richter and Furubotn (2003)).

of minimum wages and impose nondiscriminatory contract offers for all workers.[4] Workers' outside options like unemployment benefits are denoted by U and are assumed smaller than the minimum fixed wage, M. The principal is a price taker on the sales market with product price p. A linear employment contract specifies the fixed wage F and the same piece rates, r, for all workers. The principal's profit, Π, then is

$$\Pi = (p - r)(e_h + e_l) - 2 \cdot F \; . \tag{4.1}$$

Employees' earnings, ω, are

$$\omega_h = F + re_h - \frac{k}{2}e_h^2 \qquad \text{and} \qquad \omega_l = F + re_l - \frac{d}{2}e_l^2 \tag{4.2}$$

for the highly productive agent h and the less productive agent l, respectively.

The game is played finitely often. In the first x stages, the principal employs both agents; downsizing is impossible in these stages by assumption. Output is produced, learned by all parties, and sold. Profits, efforts, and earnings are assumed to be common knowledge. After the first x stages, the principal can lay off part of her labor force, i.e., downsize. More specifically, the less productive agent may be dismissed while the more productive agent remains in the firm. Although game theoretically it does not matter whether players know ex ante that downsizing is possible after x stages, behaviorally this might matter (see experimental design and discussion). In the last y stages, output is produced by the agents still employed. We now solve the downsizing game by backward induction.

■ **Effort decisions:** From (4.2) one derives the agents' optimal effort choices as $e_h^* = r/k$ and $e_l^* = r/d$ for the highly and less productive agent, respectively.

Agents obviously want to be employed since $M > U$.[5] Inserting the optimal efforts into the principal's profit function yields

$$\Pi = (p - r) \cdot (\frac{r}{k} + \frac{r}{d}) - 2 \cdot F \; . \tag{4.3}$$

[4] Collective wage agreements or strict anti-discrimination laws justify these assumptions. We do not model labor market competition (see Berninghaus et al. (2007, 2009)).

[5] To avoid further complexity, we refrain from giving agents the option to quit.

■ **Contracts before downsizing:** Denoting the contract offer before the downsizing decision by $(\tilde{F}, \tilde{r})$, the principal maximizes (4.3) subject to the minimum wage constraint (MWC1) $\tilde{F} \geq M$, which is obviously binding in optimum.

The optimal piece rate and the resulting effort levels are $\tilde{r}^* = p/2$, $e_h^* = p/(2k)$, and $e_l^* = p/(2d)$.[6] Both workers earn more than the outside option with the less productive worker earning less than the highly productive one:

$$U \;<\; \tilde{\omega_l} = M + \frac{p^2}{8d} \;<\; \tilde{\omega_h} = M + \frac{p^2}{8k} \qquad \text{due to } k < d \; .$$

The principal's profit,

$$\tilde{\Pi} = \frac{d+k}{4dk} p^2 - 2 \cdot M \; ,$$

is positive for

$$M \leq B_1 := \frac{d+k}{8dk} p^2 \; .$$

■ **Contracts after downsizing:** After laying off the less productive worker, the principal's contract offer, denoted by $(\hat{F}, \hat{r})$, maximizes

$$(p - \hat{r}) \cdot \frac{\hat{r}}{k} - \hat{F}$$

subject to the minimum wage constraint (MWC2) $\hat{F} \geq M$. The optimal piece rate and effort level are $\tilde{r}^* = p/2$ and $e_h^* = p/(2k)$. Only the highly productive worker earns more than the outside option:

$$\hat{\omega_l} = U \;<\; \hat{\omega_h} = M + \frac{p^2}{8k} \; .$$

The principal's profit,

$$\hat{\Pi} = \frac{p^2}{4k} - M \; ,$$

is positive for $M \leq B_2$ with $B_2 := p^2/(4k)$.

■ **Downsizing** is profitable for the principal if

[6] Non-negativity constraints and second-order conditions are fulfilled.

$$\hat{\Pi} - \tilde{\Pi} > 0 \qquad \text{or} \qquad -\frac{p^2}{4d} + M > 0 \,.$$

This holds true for $d \to \infty$ or, more generally, for $M > B_3$ with $B_3 := p^2/(4d)$. It is easy to show that $B_3 < B_1 < B_2$ always holds.[7] Thus, to guarantee that profits are nonnegative and downsizing is profitable, it suffices to impose $B_3 < M < B_1 < B_2$ for the minimum wage. By varying M between B_1 and B_3, the downsizing profitability is influenced. We will do this for our experimental treatments.

4.3 Experimental design

■ **Basic Design:** The experiment implemented the principal-agent model. In the beginning, three participants interacted: one principal (P–participant), one highly productive agent (H-participant), and one less productive agent (L-participant). Eight such triplets of participants formed a session with altogether 24 participants. We performed two sessions of each treatment.

The experiment consisted of $x = 2$ rounds without the possibility to downsize and $y = 2$ rounds afterwards. We refer to these four rounds as *first phase*. During the first phase, participants did not know that a perfect stranger repetition, i.e., a repetition in which no participant meets the same participants again, of the same four rounds (*second phase*) would be played afterwards. They were told, however, that another experiment would follow and that they would definitively not interact with the same participants again.

We distinguished between the following treatments: In the **announced downsizing, high incentive–treatment (AH)** all participants knew from the beginning that after the first $x = 2$ rounds the principal could downsize and that two more rounds would be played thereafter. Furthermore, the principal's theoretical profit gain from downsizing was rather large.

The only difference in the **announced downsizing, low incentive–treatment (AL)** was that the profit increase from downsizing was rather small.

In the **unannounced downsizing, high incentive–treatment (UH)** participants were not told ex ante that downsizing would be possible after two rounds,

[7] The distances between the boundaries increase with increasing d and vanish for $d \to k$.

i.e., they played the first two rounds without a hint of the subsequent downsizing opportunity. The hypothetical profit gain was the same as in AH. The first phase of this experiment was used as a separate treatment while the data of the repetition were pooled with treatment AH (see section 4.5). Since our predictions did not concern the interaction of announcing downsizing and the size of incentives and thus can be tested by these three treatments, the fourth one with unannounced downsizing and low incentives has been neglected.

After the first phase, participants in treatments AH and AL were told that one repetition of all four rounds of their corresponding treatment would be played (in a perfect stranger design). We denote the first phase of treatment AH with AH(1stphase), the second phase with AH(2ndphase), and so on. Since participants in the UH–treatment would anyhow anticipate the downsizing opportunity in the second phase, we also announced the downsizing option to them in UH(2ndphase).

■ **Treatment parameters:** We constantly set $d = 12$, $k = 2$, $U = 15$, $p = 24$, in experimental currency units (ECU).

In treatments **AH and UH**, we furthermore set $M = 24$. For the sake of readability, we denote the timing (before (B) or after downsizing (A)) with an index on the lower right and give indexes only if indispensable.

Game-theoretically, the optimal contracts before downsizing consist of $F_B = 24$ and $r_B = 12$ in both treatments. Optimal effort levels are $e_h^* = 6$ for the highly productive agent, h, and $e_l^* = 1$ for the less productive one, l. Payoffs are $\omega_{l,B} = 30$ for worker l, $\omega_{h,B} = 60$ for worker h, and $\Pi_B = 36$ for the principal. The principal should downsize and offer the same contract to the remaining agent. Payoffs are then $\omega_{l,A} = 15$, $\omega_{h,A} = 60$, and $\Pi_A = 48$. Thus, principal P can hypothetically increase the profit by about 33.3 %. Note that the principal's absolute gain is smaller than the less productive agent's loss, and welfare defined as payoff sums thus decreases even when assuming the unemployment benefit to be, miraculously, cost neutral. Efficiency thus rules out downsizing and suggests that agents double the effort levels to $e_l^+ = p/d$ and $e_h^+ = p/k$, respectively.[8]

In treatments **AL**, we set $M = 16$. The optimal contract before downsizing

[8]Here, our study also fundamentally differs from Fischer et al. (2008), where social welfare is maximized with downsizing.

consists of $F_B = 16$ and, again, $r_B = 12$. Optimal efforts are unaffected. Payoffs are lower for the workers ($\omega_{l,B} = 22$ and $\omega_{h,B} = 52$) and higher for the principal ($\Pi_B = 52$) due to the lower minimum wage. Nevertheless, the principal should downsize and offer the same contract to the highly productive agent, who invests the same effort as before. Payoffs are then $\omega_{l,A} = 15$, $\omega_{h,A} = 52$, and $\Pi_A = 56$. This means that principal P can increase the profit by about 7.7 %. Again, the agent's loss more than outweighs the principal's profit increase so that conservatively estimated welfare decreases.

We will refer to all of these results for treatments AH, UH, and AL as *benchmark predictions* from now on.

■ **Miscellaneous:** All participants received a fixed fee of $FF = 90$ ECU in each of the two phases of the experiment. We split this amount into two parts, FF_1 and FF_2 (45 ECU for rounds 1 and 2, 45 ECU for rounds 3 and 4), in the first phase of the UH–treatment. The experiment was programmed and conducted with z-Tree (see Fischbacher (2007)).

Contract offers and effort choices were restricted to reasonable intervals.[9] Furthermore, employees were prohibited from choosing effort levels that would result in negative round earnings for the given contract. Principals were asked to make conjectures about the effort choices of their agents. Their contract offer in combination with these conjectures about effort choices was not allowed to imply negative expected payoffs.[10] These restrictions could be checked by participants with a calculator integrated into the software. Each participant could use this device up to two minutes each round to calculate all resulting payoffs from any combination of F, r, e_i, and e_j.

Nevertheless principals' earnings could be negative due to overestimated efforts. The losses, if occurring, were subtracted from the other rounds' earnings and the fixed fee. We informed subjects that aggregate losses had to be paid out of pocket or by administrative work (other experiments excluded participants whose aggregate

[9] Principals were restricted to fixed wage offers F with $24 \leq F \leq 40$ in treatments AH and UH, and $16 \leq F \leq 40$ in AL. For piece-rate offers we demanded $0 \leq r \leq 20$, for efforts $0 \leq e \leq 10$. One decimal point was allowed.

[10] Other experimental studies also use techniques to avoid financial suicide of participants (see, e.g., Falk et al. (2008)).

payoff approached zero, e.g., Fehr and Schmidt (2007)). We stressed that this was very unlikely to happen. In fact, only once a moderate overall loss did occur. Representative instructions are given in Appendix A.3.

Two sessions of each treatment were played. Thus we employed 48 participants per treatment and 144 participants altogether. They were recruited, using the software ORSEE (Greiner (2004)). Before the experiment, participants had to answer a few control questions (at their computer terminal) that checked their proper understanding of the instructions. All sessions were conducted at the computer laboratory of the Max Planck Institute, Jena, in July 2008. Participants were students. Experimental sessions lasted about 100 minutes. Average earnings were € 14.55 (about € 8.70 per hour).

4.4 Hypotheses

Although the benchmarks always predict downsizing, we expect layoffs to occur less often when the principal's gain from downsizing is smaller. Many experiments have shown that other-regarding preferences play a role in human behavior.[11] Similarly, P–participants in our downsizing experiment may face a trade-off between interest for own and other participants' payoffs. Since material gains from downsizing are larger in the AH–treatment, we expect downsizing in this case more often, but not always. The finding of Charness and Rabin (2002) that efficiency concerns might be important only supports this conjecture.

Hypothesis I: *There is less downsizing in the AL-treatment than in the AH-treatment. Downsizing occurs in AH as often as in UH. There are cases without downsizing in all treatments.*

For contract offers, motives like inequity aversion, altruism, or fairness (see, e.g., Rabin (1993) or Bolton and Ockenfels (2000)) might induce at least some P–participants to offer better terms than predicted. P–participants may try to inspire higher, efficiency enhancing effort levels by increasing fixed wages and/or the piece-

[11] See Davis and Holt (1993) or Kagel and Roth (1995) for comprehensive discussions of dictator and public good experiments and, e.g., Andreoni and Miller (2002), Charness and Rabin (2002), or Fehr and Fischbacher (2003) on human altruism and social preferences.

rate offers. In a broader sense, our experiment could also be perceived as a trust game (see Cox (2004) or Kirstein and Bleich (2006)). Since too low piece-rate offers harm principal and agents, we do not expect them. In particular, contract offers significantly above the optimum could be perceived as a kind action of the principal and trigger reciprocity, i.e., effort levels above the individual benchmark predictions (Brandts and Charness (2004), Dufwenberg and Kirchsteiger (2004)).

Hypothesis II: *Piece-rate offers are at or slightly above their benchmark level. Fixed wage offers are higher than the minimum fixed wage.*

Announcing the downsizing opportunity early could affect the effort levels of less productive workers. In the AH-treatment, L–participants know that they can be fired and might exert more effort before this downsizing decision than in UH-treatment. We expect probation period effects again, i.e., that L–participants who might be fired and are aware of this should spend more effort before the downsizing decision than afterwards. It is also interesting how the remaining agents react to witnessing layoffs. They might perceive the firing of other agents as an unkind action of their principal and react reciprocally – here by lowering their effort level.

Hypothesis III: *a) Effort levels follow the pattern of a probation period effect for less productive workers anticipating downsizing. Round 1 and 2 effort levels of L–participants in the UH-treatment are lower than in the AH-treatment.*

b) Effort levels of some H–participants decrease after witnessing the firing of L–participants.

Overall, these considerations should, on average, lead to payoffs slightly above the benchmark predictions.

Hypothesis IV: *Payoffs and welfare slightly exceed their benchmark levels.*

4.5 Experimental results

In each session, 8 triplets played the game twice, i.e., the two sessions combined supplied us with 16 triplets of average data before and after downsizing. For treatment AH, e.g., we used the averages of the two rounds before downsizing as independent observations (observations: AH(1_B), AH(2_B), ..., AH(16_B)). The averages of the two rounds after downsizing were used as mutually independent observations (observations: AH(1_A), AH(2_A), ..., AH(16_A)) that depend on observations AH(1_B) - AH(16_B). Since we used a perfect stranger design, we pooled the data of the *first* and *second phases* of treatments AH and AL. Specifically, we pooled AH(1stphase) with AH(2ndphase) and AL(1stphase) with AL2(2ndphase), respectively.[12] We denote the pooled data sets by AH(32) and AL(32).

In the repetition of treatment UH, participants were probably not unaware of the downsizing opportunity. UH(1stphase) thus is a single treatment with 16 observations before and after downsizing (named UH(16)) and serves as a small-scale check of differences to treatment AH. The perfect stranger design and essentially the same procedures and instructions as in AH(2ndphase)[13] suggest to pool the data of UH(2ndphase) with AH(32). We checked this and found it confirmed by the data before downsizing, and only violated for the piece-rate offer after downsizing[14], a small lack of congruency probably due to the small number of observations in UH(2ndphase). We decided to pool the data before and after downsizing (denoted with AH(48)) and only mention the minor differences to AH(32) in footnotes.

4.5.1 Downsizing decisions

Table 4.1 summarizes the absolute and relative frequency of firms that lay off the less productive worker after the first two rounds. Downsizing occurs in all three treatments, but significantly differs from the theoretical prediction that all firms downsize (binomial tests for all three treatments: $p < .001$).

In treatment AH, 72.9 % of firms (35 of 48 firms) fire their less productive workers after two rounds and even 78.1 % do this in the AL–treatment (25 of 32 firms), although in AH theoretical gains from downsizing are with 33.3 % larger than in AL with 7.7 %. Factual gains are also higher in AH than in AL (see Section 4.5.4

[12] Checking this procedure via two sample Mann-Whitney U-tests for the most important variables, we found only one significant difference for low productive workers' efforts in AL ($p = .043$). All other tests yielded much higher p-values, data on request.

[13] They differed only in a few words in one line. For participants in AH they stated: *"again, you will receive your **participation fee**"* FF. For participants in UH they stated: *"again, you will receive both your **participation fees** (45 ECU each)"* $FF_1 + FF_2 = FF$.

[14] Mann-Whitney U-tests: 32 and 16 observations, respectively: F_W ($p = .414$), r_W ($p = .265$), $e_{h,W}$ ($p = .519$), $e_{l,W}$ ($p = .788$), F_A ($p = .516$), r_A ($p = .028$), $e_{h,W}$ ($p = .056$).

	Downsizing		No downsizing	
Treatment	Abs. freq.	Rel. freq	Abs. freq.	Rel. freq
AH	35	72.9 %	13	27.1 %
AL	25	78.1 %	7	21.9 %
UH	13	81.2 %	3	18.8 %

Table 4.1: All treatments, downsizing decisions

below). The difference in downsizing frequencies between these two treatments is insignificant, however, according to Fisher's Exact or χ^2-tests.[15] The difference between the fractions of downsizing firms in treatments AH (72.9 %) and UH (81.2 %) is also insignificant.[16]

In conclusion, *hypothesis I* predicting that the size of theoretical gains are relevant for the frequency of downsizing is clearly not in line with our findings.

Result 1: *There are only negligible differences in downsizing frequencies between all three treatments.*

4.5.2 Treatment AH

Table 4.2 describes the decisions and resulting payoffs for all participants in AH(48) before and after the downsizing decision. It distinguishes between a) all 48 observations (abbreviated and indexed *all*, if inevitable), b) firms that did downsize (*firms D*, 35 obs.), and c) firms that did not downsize (*firms ND*, 13 obs.). In cases in which (some) observations are missing, e.g., for the effort e_l of a fired worker, table cells are left empty. We will not index the treatment when it is obvious.

As expected, fixed wage offers are slightly larger than the minimum wage of 24 before and after downsizing (ranging from 24.81 to 25.79), but do not differ greatly between periods or subgroups. The difference between the fixed wages and the minimum wage is about 5.5 percent and thus about as high as the comparable

[15] The p-values for Fisher's Exact tests/standard χ^2-tests/χ^2-tests with Yates-correction are: $p = .793/.793/.792$

[16] Same tests as before: $p = .740/.740/.739$. All results stay qualitatively the same when using AH(32) instead of AH(48). In AH(32), e.g., 71.9 % of the firms chose to downsize instead of 72.9 % in AH(48).

	a) All		b) Only firms D		c) Only firms ND	
	Before	After	Before	After	Before	After
F	25.22	25.64	25.37	25.79	24.81	25.23
r	8.42	8.06	7.73	7.46	10.27	9.67
e_h	4.40	4.17	4.06	3.79	5.32	5.19
e_l	.92		.81		1.23	1.03
Π	23.24	30.09	17.72	28.90	38.12	33.31
ω_h	46.04	45.45	44.28	43.84	50.79	49.77
ω_l	27.81	18.65	28.04	15.00	27.19	28.47

Table 4.2: Treatment AH, main results

relation in Chapter 2 that solely deals with minimum wages.

Piece-rate offers, on the other hand, are significantly smaller than their benchmark level of $r = 12$.[17] This is very surprising, since calculators were in heavy use.[18] The average piece rate offered over all firms before downsizing is only $r_B = 8.42$ and remains at about the same low level, $r_A = 8.06$. Distinguishing between firms D and ND reveals that the former, on average, offer lower piece rates before downsizing, $r_{B,D} = 7.73$, than the others, $r_{B,ND} = 10.27$. After downsizing, firms D offer $r_{A,D} = 7.46$, firms ND $r_{A,ND} = 9.67$. Due to the much larger variance among firms D (std. dev.: 4.38 and 1.98), these differences are slightly insignificant on the 5%-level when performing Mann-Whitney U-tests and robust rank-order tests *(U-tests (rro tests))*.[19] In summary, *hypothesis II* is partly confirmed, and partly rejected.

Result 2.AH: *Offered fixed wages slightly exceed the benchmark predictions. Piece-*

[17] One sample t tests for benchmarks levels for a) the group of all firms and b) downsizing firms deliver the following p-values before (after) downsizing: a) $p < .001(.001)$, b) $p < .001(.001)$. The test results for a Wilcoxon paired sample test against the benchmark level for c) firms ND are: c) $p = .006(.004)$

[18] For example, principal participants in this treatment used the calculator for an average of 95 seconds in <u>each</u> of the 4+4=8 rounds they played.

[19] Before downsizing: U-tests (rro tests): $p = .087(.060)$, after downsizing: $p = .063(.052)$. The only minor difference when using AH(32) is that the difference in piece rates after downsizing becomes slightly significant. Data on request.

rate offers are below optimum throughout.

Due to the low piece rates, comparing efforts to benchmarks is not illuminative at all. Thus, Table 4.3 presents data for relative deviations, named *reldev*. We define *reldev* as the quotient between the absolute deviations (measured as the difference between the chosen and the optimal effort level) and the optimal effort. We focus on relative instead of absolute deviations, since the latter have some shortcomings that can be easily demonstrated by an example. For an offered piece rate $r = 15$, for example, an absolute deviation of $x = 0.15$ decreases a less productive worker's payoff by about 1 % and increases the principal's profit by about 12 %. If $r = 6$, however, the same absolute deviation is costlier for the worker (9 %) and more beneficial for the employer (30 %). It seems odd to treat these equal sized deviations equally, since the latter should be perceived as more generous. Using the relative deviation circumvents this problem.[20] Now, values larger (smaller) than zero indicate efforts above (below) the level predicted by models with payoff-maximizing agents.

	a) All		b) Only firms D		c) Only firms ND	
	Before	After	Before	After	Before	After
$reldev_h$	.11	.04	.13	.02	.08	.08
$reldev_l$	.62		.62		.64	.45

Table 4.3: Treatment AH, effort deviations

From Tables 4.2 and 4.3 it follows that for highly productive workers deviations are only slightly above 0. We do not observe a sharp decrease in effort levels of highly productive workers who witness layoffs. In fact, the moderate decrease of $e_{B,D} = 4.06$ to $e_{A,D} = 3.79$ is insignificant (dependent sample t test: $p = .273$) and accompanied by lower piece-rate offers.

Less productive agents tend to invest much more effort than is optimal. Relative deviations lie between .45 and .64 here and are all (almost) significantly different

[20] We thereby accept the very few undefined cases in which the piece-rate offer was $r = 0$. When only one of the two values that were used to compute averages was missing, we used the other one only.

from 0.[21] Although contract offers do not change much after downsizing among firms ND, the average effort level of workers before downsizing, $e_{B,ND} = 1.23$, is significantly higher than the average after downsizing, $e_{A,ND} = 1.03$ (Wilcoxon paired sample test, $p = .043$).[22] This confirms the probation period effect predicted in *hypothesis III*. The difference in the relative deviation between firms D and ND, $reldev_{B,D} = .62$ and $reldev_{B,ND} = .64$ (U-test: $p = .063$), weakly indicates that workers who tend to be more generous, i.e. who spend comparatively more effort, are not fired.

Result 3.AH: *Highly productive workers behave opportunistically. There is no evidence that they react to witnessing layoffs. Less productive workers tend to spend more effort than optimal, especially during probation periods.*

P–participants' average payoffs are lower than the benchmark levels, what is not surprising considering the low piece rates. There is a significant increase of about 30 % in the principals' average payoff from $\Pi_{B,all} = 23.24$ before downsizing to $\Pi_{A,all} = 30.09$ (dependent sample t test: $p = .011$). The same pattern holds when limiting the analysis to firms D where the increase from $\Pi_{B,D} = 17.72$ to $\Pi_{A,D} = 28.90$ is even larger with 63 %.[23] Firms ND, on the other hand, suffer a little payoff loss from $\Pi_{B,ND} = 38.12$ to $\Pi_{A,ND} = 33.31$.[24] Overall, this emphasizes that not only theoretical, but also factual gains render downsizing profitable. Firms D earned only $\Pi_{B,D} = 17.72$ on average before downsizing while firms ND earned more than twice as much with $\Pi_{B,ND} = 38.12$. This difference is significant according to an U-test (rro test) ($p = .001(.002)$) and seems to imply that principals' payoffs determine the downsizing decision. We will present further evidence supporting this conjecture at the end of this section.

Due to the low piece rates, payoffs of H–participants are smaller than their benchmark of 60 for all groups, but are quite invariant to the firing decision. Less productive workers, in total, suffer an income loss after downsizing: their average earnings are $\omega_B = 27.81$ before and $\omega_A = 18.65$ after downsizing due to the firings by firms

[21] One sample t tests: $p = .009$ for $reldev_{B,all}$, $p = .051$ for $reldev_{B,D}$; Wilcoxon paired sample tests: $p < .001$ for $reldev_{B,ND}$, $p = .004$ for $reldev_{A,ND}$.

[22] This result remains when restricting to AH(32) ($p = .047$). Minor changes of p-values occur for deviations. Data on request.

[23] Dependent sample t test: $p < .001$

[24] Wilcoxon paired sample test: $p = .635$

D where average earnings decrease from $\omega_{B,D} = 28.04$ before downsizing to the unemployment benefit of 15 afterwards while in firms ND, earnings do not change much ($\omega_{B,ND} = 27.19$ and $\omega_{A,ND} = 28.47$). Overall, *hypothesis IV* is rejected.

Result 4.AH: *Principals' payoffs are lower than predicted and increase after downsizing. Firms ND earn much more before downsizing than firms D. Less productive workers suffer an income loss when being fired.*

Let us shortly deal with the "cheap talk"-conjectures of agents' efforts, we asked the principals to give. We define the principal's relative conjecture accuracies for both workers' efforts as the quotients of optimum effort and conjectured effort and denote them with $relconj_h$ and $relconj_l$, respectively. Values smaller 1 suggest that principals expect higher efforts than optimal. In our experiment, the averages for all firms are below 1. In particular, we observe $relconj_{h,B} = .87$, $relconj_{h,A} = .83$, and $relconj_{l,B} = .57$ which means that principals expect more effort than optimal. Principals seem to overestimate the less productive workers' efforts ($relconj_{l,B} = .57$) comparatively more than those of high productive ones ($relconj_{h,B} = .87$). Consequentially, a one sample t test comparing the difference $conj_d := relconj_{h,B} - relconj_{l,B}$ (average: $conj_d \approx .30$) with 0 is highly significant ($p < .001$).

To confirm that a principal's downsizing decision is predominantly influenced by her own profit, we perform a simple logistic regression. It uses a forward algorithm to check which of all explanatory variables before downsizing influence the binary dependent dummy variable *Down* taking the value 1 if a firm downsizes and 0 otherwise. This algorithm uses score tests to decide which variable to include next as well as likelihood ratio test to evaluate whether the inclusion improves the model's explanatory power significantly. Formally, our model with j variables can be described by

$$Z_i = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \sum_j \beta_j x_j$$

with Z_i as the latent variable and p_i determined by the logistic function.

A selection of relevant tables is given in Appendix C.2. The algorithm stops after step 1, including only the firm's payoff Π as explanatory variable.[25] The parameter

[25]We thus need not be concerned about multicollinearity.

estimates are $\beta_0 = 3.245$ for the constant and $\beta_1 = -.075$ for the coefficient of Π. This means that the probability of firing the less productive employee is about 92.4 % when the principal's profit is $\Pi = 10$, about 63.1% at the theoretical profit before downsizing, $\Pi = 36$, and only about 40.9 % when his profit reaches $\Pi = 48$. Nagelkerke's R^2 for our regression is $R^2 = .281$.[26]

Result 5.AH: *Firms tend to overestimate effort levels, especially those of less productive workers. Firing decisions are strongly influenced by principals' profits.*

4.5.3 Treatment AL

In treatment AL, we have 32 observations in total, 25 for firms D and 7 for firms ND. Table 4.4 gives an overview of decisions and payoffs.

	a) All		b) Only firms D		c) Only firms ND	
	Before	After	Before	After	Before	After
F	18.47	17.45	18.50	17.57	18.36	17.00
r	8.96	9.36	8.83	9.61	9.43	8.46
e_h	4.26	4.34	4.13	4.40	4.74	4.14
e_l	.87		.88		.84	.80
Π	32.82	36.56	30.20	36.38	42.16	37.21
ω_h	40.03	42.35	39.62	43.83	41.52	37.07
ω_l	21.66	16.05	21.52	15.00	22.20	19.81

Table 4.4: Treatment AL, main results

Fixed wage offers before downsizing are again higher than their benchmark $F = 16$, ranging from 18.36 to 18.50 and decline by about 1 ECU after downsizing in all groups. As in treatment AH, piece rate offers are significantly below the optimum of $r = 12$ in all groups.[27] The differences between piece-rate offers of firms D and ND and before and after downsizing are negligible.

[26] Results are qualitatively the same for AH(32).

[27] One sample t test for a) the group of all firms before (after) downsizing: $p =< .001(.001)$. Test results for Wilcoxon paired sample tests for b) only firms D and c) firms ND are: b) $p = .001(.014)$, c) p =.016(.016)

Result 2.AL: *Offered fixed wages are above their benchmark. Piece-rate offers are much below optimum throughout.*

Again, H–participants do not react to layoffs of less productive workers. On the contrary, their effort is higher after downsizing, $e_{A,D} = 4.40$, than before, $e_{B,D} = 4.13$ – probably due to slightly higher piece-rates. Relative deviations, presented in Table 4.5, suggest that H–participants spend a little less effort than optimal. L–participants spend a little more effort than is optimal and do not indicate a probation period effect. Relative deviations among all firms are $reldev_{l,B} = .26$ and are almost significantly different from 0 (two-sided t test: $p = .058$).

	a) All		b) Only firms D		c) Only firms ND	
	Before	After	Before	After	Before	After
$reldev_h$	−.07	−.08	−.09	−.10	.04	−.03
$reldev_l$	.26		.31		.10	.11

Table 4.5: Treatment AL, effort deviations

Result 3.AL: *Highly productive workers behave rather opportunistically, spending less effort than optimal. There is no evidence that they react to witnessing layoffs. Less productive workers tend to spend more effort than optimal; there is no indication of a probation period effect.*

Firms' profits in treatment AL follow a pattern similar to the AH–treatment, i.e., they increase after firing the low productive worker. However, when performing a logistic regression analogous to that of the preceding subsection, the algorithm stops at the null model, not including any explanatory variable at all (see Appendix C.2 for details). Highly productive workers earn less than their benchmark and about the same before and after downsizing.

In treatment AL, firms slightly underestimate the effort choices of highly productive workers ($relconj_{h,B} = 1.16$ and $relconj_{l,B} = 1.27$), but they still overestimate the efforts of less productive workers ($relconj_{l,B} = .75$). Again, a one sample t test using the difference $conj_d$ (average: $conj_d \approx .41$) to check whether estimation accuracies differ, confirms that they, indeed, do ($p < .001$).

Result 5.AL: *Downsizing firms earn more after layoffs and less than firms ND before downsizing. Firms tend to underestimate effort levels of highly productive workers and overestimate those of less productive workers.*

4.5.4 Treatments AH and AL

When comparing the results of treatments AH and AL, we mainly concentrate on aggregate firm level. Since different minimum wages apply, fixed wage offers are, of course, lower in treatment AL than in treatment AH before downsizing ($F_{AH,B} = 25.22$ and $F_{AL,B} = 18.47$) and after downsizing ($F_{AH,A} = 25.64$ and $F_{AL,A} = 17.45$). Piece-rate offers, by contrast, are quite similar before downsizing: $r_{AH,B} = 8.42$ and $r_{AH,B} = 8.96$. A two-sided Welch-Satterthwaite t test (*WS test*) comparing them is insignificant ($p = .519$). The same is true after downsizing where piece rates are $r_{AH,A} = 8.06$ and $r_{AL,A} = 9.36$, respectively (WS test: $p = .147$). In summary, in both treatments fixed wages are a little higher than the feasible minimum while piece-rates in both treatments are too low.

Let us now take a deeper look at the differences in efforts and effort deviations for treatments AH and AL to answer the question why the fraction of downsizing firms in AH is not higher than in AL. Here, it is interesting to distinguish between subgroups again: While there are only small differences for firms D, the effort levels of less productive workers in firms ND are $e_{l,AH,B} = 1.23$ in treatment AH, but only $e_{l,AL,B} = .84$ in AL (U-test (rro test): $p = .015$ (.010)), although piece-rate offers do not differ much; they are $r_{AH,B} = 10.27$ in treatment AH and $r_{AH,B} = 9.43$ in AL and this difference is insignificant (U-test: $p = .263$).

This suggests that L–participants who are not fired, later on, in treatment AH might be concerned about future layoffs and thus spend relatively more effort than those in treatment AL to prevent their principals from firing them. This view is supported by the finding that relative deviations among the same worker group are $reldev_{l,AH,B} = .64$ in AH, but only $reldev_{l,AL,B} = .10$ in AL, i.e., low productive workers deviate comparatively more from optimum effort and are thereby more generous in treatment AH. The descriptive finding is confirmed by an U-test (rro test), since the reported difference is significant with a p-value of $p = .037$ (.027). Interestingly, across all firms even high productive workers are slightly more generous in treatment AH with $reldev_{h,AH,B} = .11$ than in AL with $reldev_{l,AH,B} = -.07$ (WS test: $p = .040$).

We think that these results help to understand why firms do not downsize more often in treatment AH than in AL, especially when additionally considering that – despite the more generous effort levels – not only theoretical gains, but also the factual gains from downsizing are larger in AH than in AL. It seems as if workers intuitively evaluate their weaker position in AH correctly and principals are appeased by the rather generous effort decisions of low (and high) productive workers in this treatment.

Payoffs of workers are about the same in both treatments except for the rent reallocation due to the lower fixed wages.[28]

Result 6: *Treatments AH and AL differ in fixed wages and thus in payoffs: while principals earn more in AL, agents earn less. Piece-rate offers in both treatments hardly differ. Before downsizing, especially the less productive workers who are not fired spend relatively more effort in treatment AH than those in treatment AL.*

4.5.5 Treatment(s) UH (and AH)

In contrast to treatment AH, participants in UH were not aware of the downsizing opportunity. We performed UH as a first small check to get an impression whether agents' behavior differs when the firing threat hangs over them like a Sword of Damocles as in treatment AH. Due to the small sample size, we only list the main results of UH and state important differences to AH, here. Tables 4.6 and 4.7 with three firms ND, 13 firms D and 16 firms in total are constructed analogously to the preceding tables.

Fixed wage offers are above optimum, but do not differ before and after downsizing. Piece-rate offers are always far below the optimum. Effort levels are relatively the same across groups of firms and do not differ before and after downsizing. There is no probation period effect (less productive workers' effort is almost identical before, $e_{B,ND} = .67$, and after downsizing, $e_{A,ND} = .68$, among firms ND). Deviations

[28]The only noteworthy difference when using AH(32) instead of AH(48) is that after downsizing piece rates are now significantly larger in AL than in AH (WS test: $p = .032$).

	a) All		b) Only firms D		c) Only firms ND	
	Before	After	Before	After	Before	After
F	26.52	25.71	26.79	25.95	25.33	24.67
r	7.98	8.93	7.98	9.11	8.00	8.17
e_h	4.46	4.48	4.37	4.47	4.83	4.50
e_l	.79		.82		.67	.68
Π	22.65	33.31	23.14	35.02	20.53	25.88
ω_h	44.74	48.13	44.28	49.24	46.75	43.33
ω_l	28.97	17.34	28.95	15.00	29.05	27.47

Table 4.6: Treatment UH, main results

are comparable to treatment AH, except those of less productive workers among firms ND (see Table 4.7).

	a) All		b) Only firms D		c) Only firms ND	
	Before	After	Before	After	Before	After
$reldev_h$	.15	.00	.08	$-.02$	.46	.08
$reldev_l$	.55		.75		$-.29$	$-.19$

Table 4.7: Treatment UH, effort deviations

Just as in treatment AH, payoffs of all groups are lower than predicted, although the difference is insignificant for less productive employees. The principals' average payoff gain for all firms from before to after downsizing is large, but not significant on the 5%-level (Wilcoxon paired sample test, $p = .093$). H–participants earn almost the same in all groups. Less productive employees suffer a significant income loss when being fired and earn about the same before and after downsizing otherwise.[29] Relative conjecture accuracy is similar to treatment AH: principals overestimate highly productive workers' efforts ($relconj_{h,B} = .76$ and $relconj_{h,A} = .86$) as well as the effort of L–participants ($relconj_{l,B} = .54$), but conjectures are still

[29]We did not calculate a logistic regression because of the small sample size.

more accurate for highly productive workers.[30]

When comparing treatments AH and UH, we first checked all contract offers and payoffs (a complete list of tests is given in Appendix C.1) and found no significant differences what could have been expected.

One major difference between AH and UH is the rather large, but insignificant difference between effort levels of less productive workers (in firms ND) in treatment AH, $e_{AH,B} = 1.23$, and in treatment UH, $e_{AH,B} = .67$ (U-test (rro test): $p = .179(.203)$), partly due to lower piece-rate offers. This, however, cannot account for the large difference between relative deviations among firms ND ($reldev_{AH,B} = .64$ vs. $reldev_{UH,B} = -.29$), that is significant, despite the small sample size of firms ND and the conservative two-sided application of the test (U-test (rro test), $p = .041(.041)$). Apparently, L–participants who anticipate the downsizing decision (treatment AH) are more generous than those who cannot foresee being fired (treatment UH) – at least among the firms that eventually do not fire them.[31] For comparison, among firms D the relative deviations do not differ much with $reldev_{AH,B} = .62$ and $reldev_{UH,B} = .75$, respectively.[32] Analogously, the probation period effect the data suggest for treatment AH in which efforts decrease from $e_{AH,B} = 1.23$ to $e_{AH,A} = 1.03$ cannot be found in treatment UH in which effort levels hardly differ with $e_{AH,B} = .67$ and $e_{AH,A} = .68$. This confirms the respective part of *hypothesis III.a)* and we summarize:

Result 7: *The main trends in treatment UH are largely comparable to those of treatment AH, but among firms ND, L–participants are more generous in AH than in UH, especially before the downsizing decision.*

4.6 Conclusion

In this chapter, we theoretically and experimentally analyzed downsizing decisions by low and highly profitable firms in a principal-agent-setting.

Contrary to our conjectured hypotheses, firms did not downsize more often in

[30] One sample t test, difference $conj_d$ (average: $conj_d \approx .22$) against 0, $p < .001$.
[31] Again, all these results stay qualitatively the same for AH(32).
[32] U-test: $p = .790$.

treatments in which theoretical and factual profitability were rather high. This was captured by treatments AH and AL in which those profitability measures largely differed, but no significant differences in frequencies of downsizing firms were found. However, we did find some evidence that within a specific treatment the most unprofitable firms were those that eventually downsized their labor force.

Our findings also suggested that worker participants in the treatment with theoretically higher downsizing incentives were more generous, i.e., they were spending comparatively more effort than workers in the opposing treatment with rather low downsizing incentives. This might have been a way by which workers mitigated their employers and thereby prevented firings.

Additionally, the distinction between treatments AH and UH enabled us to highlight the behavioral differences resulting from anticipating possible future layoffs. If downsizing opportunities could not have been anticipated (as in treatment UH), there was no probation period effect what might explain why sometimes firms in the real world threaten to lay off (larger shares of) their workforce and withdraw from this measure later on. Overall, the workers who were not fired in our experimental treatments were spending less effort in the treatment in which they did not foresee the downsizing opportunity.

5 Summary and conclusion

With this thesis we tried to shed further light on three aspects of labor markets: minimum wage spillovers, the interplay of employment protection and bullying, and layoff policies of firms operating highly profitable and less profitable, respectively. We are confident that our theoretical and experimental insights have significantly enriched the knowledge about these three aspects of labor markets. We now want to briefly summarize the main results of each of our three main chapters and suggest some starting points for future research.

We first asked whether minimum wage spillovers occur, i.e., whether the introduction of a minimum wage not only increases the wages of workers who previously earned less than this minimum wage, but also of those who earned considerably more already. A theoretical model was presented that focused on only one of the couple of causes for minimum wage spillovers discussed in literature, namely on relative income preferences, and thereby excluded other causes like substitution effects, for example. The model was not limited to the two utility functions capturing relative income preferences often proposed in literature, but showed for a more general set of assumptions that minimum wage spillovers follow. The concentration on relative income preferences allowed us to design an experiment that was capable to exclude the same causes for minimum wage spillovers as the model did. In the experiment minimum wage spillovers occurred and a control treatment suggested that, indeed, relative income preferences caused this spillovers.

With this study, we hope to inspire future work that theoretically and experimentally investigates the interplay of relative income preferences and changes in institutional frameworks or workers' abilities. For example, it would be interesting to analyze a situation in which a specific worker's skill level changes after some periods, while the other workers' abilities stay the same.

Another straightforward extension would be to modify the deduction rule of the minimum wage such that it would be introduced slightly below or equal to the wage of the worker who occupies the middle position in the earnings hierarchy. This might lead to new theoretical results and differences in participants' laboratory behavior.

In Chapter 3, we investigated the interplay between employment protection legislation and the bullying of workers. Wasmer (2006) introduced and empirically tested the idea that firms react to stricter employment protection by trying to bully their unwanted employees forcing them to quit voluntarily. We connected his work with the experimental studies on sanctions by interpreting bullying as a costly sanction imposed by employers. Our main results were that the degree of employment protection and sanctions had the expected effects on workers' effort levels and voluntary resignations seemed, indeed, to be frequently triggered by bullying/sanctions. However, employers often used sanctions and firings as complements, rather than as substitutes.

Evaluating these main findings, we think that not only the relationship between bullying and employment protection is an interesting topic for future research, but also the more general question whether different kinds of disciplinary incentives are used together or concurrently.

Finally, firms' firing policies were analyzed in Chapter 4. We experimentally investigated the hypothesis that firms operating highly profitable are less likely to fire their unproductive employees than less profitable firms. Our main results were that, unexpectedly, downsizing was not more frequent in treatments in which firms' theoretical and factual gains from laying off part of their labor force were highest. We presented evidence suggesting that workers might have (intuitively) caused their employers' behavior by spending comparatively more effort when theoretical and factual gains were rather large, especially during probation periods and when they could anticipate the upcoming downsizing decisions.

We are not aware of experimental studies solely focusing on probation period effects. This topic seems interesting for future research. Our comparisons of behavioral differences between treatments with announced and unannounced downsizing rely on small samples yet. Further investigations might be fruitful.

Bibliography

Adams, J. S.: 1965, Inequity in social exchange, *in* L. Berkowitz (ed.), *Advances in Experimental Social Psychology*, Academic Press.

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

Alesina, A., Di Tella, R. and MacCulloch, R.: 2004, Inequality and happiness: Are Europeans and Americans different?, *Journal of Public Economics* **88**(9-10), 2009 – 2042.

Alpizar, F., Carlsson, F. and Johansson-Stenman, O.: 2005, How much do we care about absolute versus relative income and consumption?, *Journal of Economic Behavior and Organization* **56**, 405 – 421.

Andreoni, J.: 1995, Cooperation in public-goods experiments: Kindness or confusion, *American Economic Review* **85**(4), 891 – 904.

Andreoni, J., Harbaugh, W. and Vesterlund, L.: 2003, The carrot or the stick: Rewards, punishments, and cooperation, *American Economic Review* **93**(3), 893 – 902.

Andreoni, J. and Miller, J.: 2002, Giving according to GARP: an experimental test of the consistency of preferences for altruism, *Econometrica* **70**(2), 37 – 53.

Bauer, T. K., Kluve, J., Schaffner, S. and Schmidt, C. M.: 2008, Fiscal effects of minimum wages, *Ruhr Economic Papers (Discussion Paper Series)* **79**, 1 – 32.

Belot, M., Boone, J. and van Ours, J.: 2007, Welfare effects of employment protection, *Economica* **74**(295), 381 – 396.

Benz, M. and Meier, S.: 2008, Do people behave in experiments as in the field?– evidence from donations, *Experimental Economics* **11**(3), 268 – 281.

Berninghaus, S. K., Güth, W., Hoppe, C. and Paul, C.: 2007, International competition in hiring labour and selling output: A theoretical and experimental analysis, *in* W. Franz, J. Ramser and M. Stadler (eds), *Dynamik internationaler Märkte. Wirtschaftswissenschaftliches Seminar Ottobeuren*, Mohr Siebeck.

Berninghaus, S. K., Güth, W., Hoppe, C. and Paul, C.: 2009, Duopolistic hiring and sales competition: A theoretical and experimental analysis, *Working Paper* .

Blanchflower, D. G. and Oswald, A. J.: 2004, Well-being over time in Britain and the USA, *Journal of Public Economics* **88**(7-8), 1359 – 1386.

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

Boskin, M. J. and Sheshinski, E.: 1978, Optimal redistributive taxation when individual welfare depends upon relative income, *Quarterly Journal of Economics* **92**(4), 589 – 601.

Brandts, J. and Charness, G.: 2000, Hot vs. cold: Sequential responses and preference stability in experimental games, *Experimental Economics* **2**(3), 227 – 238.

Brandts, J. and Charness, G.: 2004, Do labour market conditions affect gift exchange? Some experimental evidence, *Economic Journal* **114**, 684 – 708.

Brandts, J. and Cooper, D. J.: 2006, A change would do you good .... an experimental study on how to overcome coordination failure in organizations, *American Economic Review* **96**(3), 669 – 693.

Brosig, J., Weimann, J. and Yang, C.-L.: 2003, The hot versus cold effect in a simple bargaining experiment, *Experimental Economics* **6**(1), 75 – 90.

Brown, G., D. A., Gardner, J., Oswald, A., J. and Qian, J.: 2008, Does wage rank affect employees' well-being?, *Industrial Relations* **47**(3), 355 – 389.

Camerer, C.: 2003, *Behavioral Game Theory: Experiments in Strategic Interactions*, Princeton University Press.

Card, D. E. and Krueger, A. B.: 1994, Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania, *American Economic Review* **84**(4), 772 – 793.

Card, D. E. and Krueger, A. B.: 1995, *Myth and Measurement: The New Economics of the Minimum Wage*, Princeton University Press.

Card, D. E. and Krueger, A. B.: 2000, Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania: Reply, *American Economic Review* **90**(5), 1397 – 1420.

Casari, M. and Vason, T. N.: 2009, The strategy method lowers measured trustworthy behavior, *Economics Letters* **103**, 157 – 159.

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

Charness, G. and Grosskopf, B.: 2001, Relative payoffs and happiness: an experimental study, *Journal of Economic Behavior and Organization* **45**, 301 – 328.

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

Clark, A. E., Diener, E., Georgellis, Y. and Lucas, R. E.: 2008, Lags and leads in life satisfaction: A test of the baseline hypothesis, *Economic Journal* **118**, F222 – F243.

Clark, A. E., Frijters, P. and Shields, M.: 2008, Relative income, happiness and utility: An explanation for the Easterlin Paradox and other puzzles, *Journal of Economic Literature* **46**(1), 95 – 144.

Clark, A. E., Kristensen, N. and Westergard-Nielsen, N.: 2009, Job satisfaction and co-worker wages: Status or signal?, *Economic Journal* **119**, 430 – 447.

Clark, A. E. and Oswald, A. J.: 1996, Satisfaction and comparison income, *Journal of Public Economics* **61**(3), 359 – 381.

Clark, A. E. and Oswald, A. J.: 1998, Comparison-concave utility and following behaviour in social and economic settings, *Journal of Public Economics* **70**(1), 133 – 155.

Cox, J. C.: 2004, How to identify trust and reciprocity, *Games and Economic Behavior* **46**, 260 – 281.

Critchlow, E. D. and Fligner, M. A.: 1991, On distribution-free multiple comparisons in the one-way analysis of variance, *Communications in Statistics - Theory and Methods* **20**(1), 127 – 139.

Davis, D. D. and Holt, C. A.: 1993, *Experimental Economics*, Princeton University Press.

de Quervain, D. J.-F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A. and Fehr, E.: 2004, The neural basis of altruistic punishment, *Science* **305**, 1254 – 1258.

Deaton, A.: 2003, Health, inequality, and economic development, *Journal of Economic Literature* **41**(1), 113 – 158.

Di Tella, R. and MacCulloch, R.: 2006, Some uses of happiness data in economics, *Journal of Economic Perspectives* **20**(1), 25 – 46.

Dickens, R. and Manning, A.: 2004, Spikes and spill-overs: The impact of the national minimum wage distribution in a low-wage sector, *Economic Journal* **114**, C95 – C101.

Die Zeit: 2009a, Mindestlöhne überall, *Die Zeit* **June, 4**, 24.

Die Zeit: 2009b, Seid umschlungen, Gegner!, *Die Zeit* **August, 20**, 3.

Diener, E., Sandvik, E., Seidlitz, L. and Diener, M.: 1993, The relationship between income and subjective well-being: Relative or absolute?, *Social Indicators Research* **28**, 195 – 223.

Diener, E., Suh, E. M., Lucas, R. E. and Smith, H. L.: 1999, Subjective well-being: Three decades of progress, *Psychological Bulletin* **125**(2), 276 – 302.

DiNardo, J., Fortin, N. M. and Lemieux, T.: 1996, Labor market institutions and the distribution of wages, 1973 - 1992: A semiparametric approach, *Econometrica* **64**(5), 1001 – 1044.

Dolado, J., Kramarz, F., Machin, S., Manning, A., Margolis, D. and Teulings, C.: 1996, The economic impact of minimum wages in Europe, *Economic Policy* **October**, 317 – 372.

Drazen, A.: 1986, Optimal minimum wage legislation, *Economic Journal* **96**(383), 774 – 784.

Duesenberry, J. S.: 1949, *Income, Saving and the Theory of Consumer Behaviour*, Harvard University Press.

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

Easterlin, R. A.: 1974, Does economic growth improve the human lot? Some empirical evidence, *in* P. A. David and M. W. Reder (eds), *Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz*, Academic Press.

Falk, A. and Fehr, E.: 2003, Why labour market experiments?, *Labour Economics* **10**, 399 – 406.

Falk, A., Fehr, E. and Zehnder, C.: 2006, Fairness perceptions and reservation wages - the behavioral effects of minimum wage laws, *Quarterly Journal of Economics* **121**(4), 1349 – 1381.

Falk, A. and Fischbacher, U.: 2006, A theory of reciprocity, *Games and Economic Behavior* **54**, 293 – 315.

Falk, A. and Gächter, S.: 2008, Experimental labour economics, *in* S. N. Durlauf and L. E. Blume (eds), *The New Palgrave Dictionary of Economics*, Palgrave Macmillan.

Falk, A., Huffman, D. and MacLeod, W. B.: 2008, Institutions and contract enforcement, *IZA Discussion Paper* **3435**.

Falk, A. and Huffmann, D.: 2007, Studying labor market institutions in the lab: Minimum wages, employment protection and workfare, *Journal of Theoretical and Institutional Economics* **163**(1), 30 – 45.

Falk, A. and Knell, M.: 2004, Choosing the Joneses: Endogeneous goals and reference standards, *Scandinavian Journal of Economics* **106**(3), 417 – 435.

FAZ: 2008a, BMW fährt Rekord ein - Aktie stottert hinterher, *Frankfurter Allgemeine Zeitung* **January, 30**, 16.

FAZ: 2008b, Nokia: Bochums Ertragskraft wird falsch dargestellt, *Frankfurter Allgemeine Zeitung* **January, 31**, 18.

FAZ: 2009, Schrille Töne im Wahlkampf, *Frankfurter Allgemeine Zeitung* **August, 18**, 1.

Fehr, E. and Fischbacher, U.: 2003, The nature of human altruism, *Nature* **425**, 785 – 791.

Fehr, E., Fischbacher, U. and Kosfeld, M.: 2005, Neuroeconomic foundations of trust and social preferences: Initial evidence, *American Economic Review* **95**(2), 346 – 351.

Fehr, E. and Gächter, S.: 2000a, Cooperation and punishment in public good experiments, *American Economic Review* **90**(4), 980 – 994.

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

Fehr, E. and Gächter, S.: 2002, Altruistic punishment in humans, *Nature* **415**, 137 – 140.

Fehr, E. and Rockenbach, B.: 2004, Human altruism: economic, neural, and evolutionary perspectives, *Current Opinion in Neurobiology* **14**, 784 – 790.

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

Fehr, E. and Schmidt, K. M.: 2006, The economics of fairness, reciprocity and altruism: Experimental evidence, *in* S. C. Kolm and J. M. Ythier (eds), *Handbook of the Economics of Giving*, Vol. 1, North Holland.

Fehr, E. and Schmidt, K. M.: 2007, Fairness and contract design, *Econometrica* **75**(1), 121 – 154.

Feltovich, N.: 2003, Nonparametric tests of differences in medians: Comparison of the Wilcoxon-Mann-Whitney and robust rank-order tests., *Experimental Economics* **6**(3), 273 – 297.

Feltovich, N.: 2005, Critical values for the robust rank-order test, *Communications in Statistics - Simulation and Computation* **34**(3), 525 – 547.

Ferrer-i-Carbonell, A.: 2005, Income and well-being: an empirical analysis of the comparison income effect, *Journal of Public Economics* **89**(5-6), 997 – 1019.

Festinger, L.: 1954, A theory of social comparison processes, *Human Relations* **7**, 117 – 140.

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

Fischer, S., Gueth, W. and Koehler, C.: 2008, Effects of profitable downsizing on collective bargaining, *Jena Economic Research Papers (Discussion Paper Series)* **2008-11**.

Fliessbach, K., Weber, B., Trautner, P., Dohmen, T., Sunde, U., Elger, C. E. and Falk, A.: 2007, Social comparison affects reward-related brain activity in the human ventral striatum, *Science* **318**, 1305 – 1308.

Fligner, M. A. and Policello, G. E.: 1981, Robust rank procedures for the Behrens-Fisher problem, *Journal of the American Statistical Association* **76**(373), 162 – 168.

Flinn, C. J.: 2006, Minimum wage effects on labor market outcomes under search, matching, and endogenous contact rates, *Econometrica* **74**(4), 1013 – 1062.

Flinn, C. J.: 2008, On-the-job search, minimum wages, and labor market outcomes in an equilibrium bargaining framework, *Discussion Paper Series, Institute for Research on Poverty* (1337-08).

Frank, R. H.: 1984a, Are workers paid their marginal products?, *American Economic Review* **74**(4), 549 – 571.

Frank, R. H.: 1984b, Interdependent preferences and the competitive wage structure, *RAND Journal of Economics* **15**(4), 510 – 520.

Frank, R. H.: 1985, *Choosing the Right Pond. Human Behavior and the Quest for Status*, Oxford University Press.

Frank, R. H. and Sunstein, C. R.: 2001, Cost-benefit analysis and relative position, *University of Chicago Law Review* **68**, 323 – 374.

Frankfurter Rundschau: 2008, BMW spart beim Stammpersonal., *Frankfurter Rundschau* **February, 28**, 18. 18.

Frey, B. and Osterloh, M.: 2002, *Successful Management by Motivation: Balancing Intrinsic and Extrinsic Incentives*, Springer-Verlag.

Galizzi, M. and Lang, A. K.: 1998, Relative wages, wage growth, and quit behavior, *Journal of Labor Economics* **16**(2), 367 – 391.

Garibaldi, P.: 1998, Job flow dynamics and firing restrictions, *European Economic Review* **42**, 245 – 272.

Gächter, S. and Fehr, E.: 2002, Fairness in the labour market: a survey of experimental results, *in* B. Friedel and M. Lehmann-Waffenschmidt (eds), *Surveys in Experimental Economics. Bargaining, Cooperation and Election Stock Markets.*, Physica.

Greiner, B.: 2004, An online recruitment system for economic experiments, *in* K. Kremer and V. Macho (eds), *Forschung und wissenschaftliches Rechnen,* *GWDG Bericht 63*, Gesellschaft für Wissenschaftliche Datenverarbeitung, Göttingen.

Gürerk, O., Irlenbusch, B. and Rockenbach, B.: 2006, The competitive advantage of sanctioning institutions, *Science* **312**, 108 – 111.

Grossman, J. B.: 1983, The impact of the minimum wage on other wages, *Journal of Human Resources* **18**, 359 – 378.

Güth, W., Schmidt, C. and Sutter, M.: 2007, Bargaining outside the lab - a newspaper experiment of a three-person ultimatum game, *Economic Journal* **117**, 449 – 469.

Güth, W., Schmittberger, R. and Schwarze, B.: 1982, An experimental analysis of ultimatum bargaining, *Journal of Economic Behavior and Organization* **3**, 367 – 388.

Guardian: 2009, Financial: Rio Tinto lays off 16,000 after profits fall by more than half, *The Guardian* **August, 21**, 27.

Hammermesh, D. S.: 1975, Interdependence in labour market, *Economica* **42**(168), 420 – 429.

Handelsblatt: 2008, Henkel baut 3000 Stellen ab und BMW greift hart durch, *Handelsblatt* **February, 28**, 14 – 17.

Harrison, G. W., List, J. A. and Towe, C.: 2007, Naturally occurring preferences and exogenous laboratory experiments: A case study of risk aversion, *Econometrica* **75**(2), 433 – 458.

Haviland, M. G.: 1990, Yates's correction for continuity and the analysis of 2x2 contingency tables, *Statistics in Medicine* **9**, 363 – 367.

Healy, P. J.: 2007, Group reputations, stereotypes, and cooperation in a repeated labor market, *American Economic Review* **97**(5), 1751 – 1773.

Herrmann, B., Thöni, C. and Gächter, S.: 2008, Antisocial punishment across societies, *Science* **319**, 1362 – 1367.

Hirschman, A. O. and Rothschild, M.: 1973, The changing tolerance for income equality in the course of economic development (Hirschman) with a mathematical appendix (Rothschild), *Quarterly Journal of Economics* **87**(4), 544 – 566.

Hopenhayn, H. and Rogerson, R.: 1993, Job turnover and policy evaluation: A general equilibrium analysis, *Journal of Political Economy* **101**(5), 915 – 938.

Ichino, A. and Riphahn, R. T.: 2005, The effect of employment protection on worker effort - a comparison of absenteeism during and after probation, *Journal of the European Economic Association* **3**(1), 120 – 143.

Irish Times: 2009a, Minimum wage levels, *Irish Times* **July, 25**, 15.

Irish Times: 2009b, Minimum wage levels may need adjustment, says Lenihan, *Irish Times* **July, 22**, 1.

Johansson-Stenman, O., Carlsson, F. and Daruvala, D.: 2002, Measuring future grandparents' preferences for equality and relative standing, *Economic Journal* **112**, 362 – 383.

Kagel, J. K. and Roth, A. E.: 1995, *Handbook of experimental economics*, Princeton: Princeton University Press.

Kahneman, D., Knetsch, J. L. and Thaler, R. H.: 1991, Anomalies: The endowment effect, loss aversion, and status quo bias, *Journal of Economic Perspectives* **5**(1), 193 – 206.

Kahneman, D. and Krueger, A. B.: 2006, Developments in the measurement of subjective well-being, *Journal of Economic Perspectives* **20**(1), 3 – 24.

Kahneman, D. and Thaler, R. H.: 2006, Anomalies. Utility maximization and experienced utility, *Journal of Economic Perspectives* **20**(1), 221 – 234.

Karlan, D. S.: 2005, Using experimental economics to measure social capital and predict financial decisions, *American Economic Review* **95**(5), 1688 – 99.

Katz, L. F. and Krueger, A. B.: 1992, The effect of the minimum wage on the fast-food industry, *Industrial and Labor Relations Review* **46**(1), 6 – 21.

Kimball, M. and Willis, R.: 2006, Utility and happiness, *Working Paper* .

Kirstein, A.: 2008, Bonus, malus, and fixed pay in principal-agent relationships with real effort, *Schmalenbach Business Review* **60**, 280 – 303.

Kirstein, A. and Bleich, S.: 2006, Screening of endogenous types: an experiment, *Working Paper* .

Kirstein, R., Kittner, M. and Schmidtchen, D.: 2000, Kündigungsschutzrecht in den USA und Deutschland: Ein Beitrag zur ökonomischen Rechtsvergleichung, *CSLE Discussion Paper* **2000-08**.

Lang, K.: 1987, Pareto improving minimum wage laws, *Economic Inquiry* **25**(1), 145 – 158.

Lee, D. S.: 1999, Wage inequality in the United States during the 1980s: Rising dispersion or failing minimum wage?, *Quarterly Journal of Economics* **114**(3), 977 – 1023.

Leigh, A.: 2003, Employment effects of minimum wages: Evidence from a quasi-experiment, *Australian Economic Review* **36**(4), 361 – 373.

Leigh, A.: 2004a, Employment effects of minimum wages: Evidence from a quasi-experiment - Erratum, *Australian Economic Review* **37**(1), 102 – 105.

Leigh, A.: 2004b, Minimum wages and employment: Reply, *Australian Economic Review* **37**(2), 173 – 179.

Levitt, S. D. and List, J. A.: 2007, What do laboratory experiments measuring social preferences reveal about the real world?, *Journal of Economic Perspectives* **21**(2), 153 – 174.

Luttmer, E. F. P.: 2005, Neighbors as negatives: Relative earnings and well-being, *Quarterly Journal of Economics* **120**(3), 932 – 1002.

Major, B., Testa, M. and Bylsma, W. H.: 1991, Responses to upward and downward social comparisons: The impact of esteem-relevance and perceived control, *in*

J. M. Suls and T. A. Wills (eds), *Social Comparison. Contemporary Theory and Research*, Lawrence Erlbaum Associates, Inc.

Manning, A.: 2003, *Monopsony in Motion. Imperfect Competition in Labor Markets*, Princeton University Press.

Marmot, M. and Bobak, M.: 2000, International comparators and poverty in health in Europe, *British Medical Journal* **321**, 1124 – 1128.

McBride, M.: 2001, Relative-income effects on subjective well-being in the cross-section, *Journal of Economic Behavior and Organization* **45**, 251 – 278.

McCabe, K. A., Rigdon, M. L. and Smith, V. L.: 2003, Positive reciprocity and intensions in trust games, *Journal of Economic Behavior and Organization* **52**, 267 – 275.

Mortensen, D. T. and Pissarides, C.: 1994, Job creation and job destruction in the theory of unemployment, *Review Of Economic Studies* **61**(3), 397 – 415.

Moser, B. K., Stevens, G. R. and Watts, C. L.: 1989, The two-sample t test versus Satterthwaite's approximate F test, *Communciations in Statistics - Theory and Methods* **18**(11), 3963 – 3975.

Neuhäuser, M.: 2002, Two-sample tests when variances are unequal, *Animal Behaviour* **63**, 823 – 825.

Neumark, D. and Postlewaite, A.: 1998, Relative income concerns and the rise in married women's employment, *Journal of Public Economics* **70**(1), 157 – 183.

Neumark, D. and Wascher, W.: 2000, Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania: Comment, *American Economic Review* **90**(5), 1362 – 1396.

New York Times: 2008, After a downsizing, how to motivate?, *New York Times* **August, 24**, 8. page 8.

New York Times: 2009a, Job losses slow, signaling momentum for a recovery, *The New York Times* **August, 8**, A1.

New York Times: 2009b, Where the jobs are, *New York Times* **July, 24**, 24.

Oxoby, R. J. and McLeish, K. N.: 2004, Sequential decision and strategy vector methods in ultimatum bargaining: evidence on the strength of other-regarding behavior, *Economics Letters* **84**, 399 – 405.

Pollak, R. A.: 1976, Interdependent preferences, *American Economic Review* **66**(3), 309 – 320.

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

Randles, R. H.: 2006, Wilcoxon signed rank test, *in* S. Kotz, C. B. Read, N. Balakrishan and B. Vadakovic (eds), *Encyclopedia of statistical sciences*, Wiley-Interscience.

Rayo, L. and Becker, G. S.: 2007, Evolutionary efficiency and happiness, *Journal of Political Economy* **115**(2), 302 – 337.

Rebitzer, J. B. and Taylor, L. J.: 1995, The consequences of minimum wage laws. Some new theoretical ideas, *Journal of Public Economics* **56**(2), 245 – 255.

Richter, R. and Furubotn, E. G.: 2003, *Neue Institutionenökonomik*, 3rd edn, Tübingen: Mohr Siebeck.

Riedl, A. and Tyran, J.-R.: 2005, Tax liability side equivalence in gift-exchange labor markets, *Journal of Public Economics* **89**(12), 2369 – 2382.

Rockenbach, B. and Milinski, M.: 2006, The efficient interaction of indirect reciprocity and costly punishment, *Nature* **444**, 718 – 723.

Ruxton, G. D.: 2006, The unequal variance t-test is an underused alternative to student's t-test and the Mann-Whitney U test, *Behavioral Ecology* **17**, 688 – 690.

Schellhaaß, H. M. and Nolte, A.: 1999, Employment protection in an institutional perspective, *Jahrbuch für Nationalökonomie und Statistik* **218**(3+4), 415 – 432.

Seale, D. A. and Rapoport, A.: 2000, Elicitation of strategy profiles in large group coordination games, *Experimental Economics* **3**(2), 153 – 179.

Senik, C.: 2004, Relativizing relative income, *Delta Working Paper (Discussion Paper Series)* **2004-17**.

Senik, C.: 2008, Ambition and jealousy: Income interactions in the "Old" Europe versus the "New" Europe and the United States, *Econometrica* **75**(299), 495 – 513.

Senik, C.: 2009, Direct evidence on income comparisons and their welfare effects, *Journal of Economic Behavior and Organization, forthcoming* .

Siegel, S. and Castellan, N. J.: 1988, *Nonparametric statistics for the behavioral sciences*, 2nd edn, New York: McGrawHill, Inc.

Solnick, S. J.: 2007, Cash and alternate methods of accounting in an experimental game, *Journal of Economic Behavior and Organization* **62**, 316 – 321.

Solnick, S. J. and Hemenway, D.: 1998, Is more always better? A survey on positional concerns, *Journal of Economic Behavior and Organization* **37**, 373 – 383.

Spiegel: 2009, Schlappes Gespenst, *Der Spiegel* **August, 24**, 20 – 23.

Stewart, M. B.: 2004, The employment effects of the national minimum wage, *Economic Journal* **114**, C110 – C116.

Stigler, G. J.: 1946, The economics of minimum wage legislation, *American Economic Review* **36**(3), 358 – 365.

Summers, L. H.: 1988, Relative wages, efficiency wages, and Keynesian unemployment, *American Economic Review* **78**(2), 383 – 388.

Teulings, C. N.: 2000, Aggregation bias in elasticities of substitution and the minimum wage paradox, *International Economic Review* **41**(2), 359 – 398.

Teulings, C. N.: 2003, The contribution of minimum wages to increasing wage inequality, *Economic Journal* **113**, 801 – 833.

The Australian: 2009, Wages freeze to protect jobs - Gillard and unions furious at minimum pay decisions, *The Australian* **July, 8**, 1.

Times: 2009, BT suspends graduate recruitment programme, *The Times* **August, 24**, 38.

Toothaker, L. E. and Newman, D.: 1994, Nonparametric competitors to the two-way ANOVA, *Journal of Educational and Behavioral Statistics* **19**(3), 237 – 273.

Tversky, A. and Kahneman, D.: 1981, The framing of decisions and the psychology of choice, *Science* **211**, 453 – 458.

van den Berg, G. J.: 2003, Multiple equilibria and minimum wages in labor markets with informational frictions and heterogeneous production technologies, *International Economic Review* **44**(4), 1337 – 1357.

Veenhoven, R.: 1991, Is happiness relative?, *Social Indicators Research* **24**, 1 – 34.

Walster, E., Walster, G. W. and Berscheid, E.: 1978, *Equity. Theory and Research*, Allyn & Bacon.

Washington Post: 2009, Some attack timing of minimum wage hike, *Washington Post* **July, 24**, A16.

Wasmer, E.: 2006, The economics of Prozac: Do employees really gain from strong employment protection?, *IZA Discussion Paper* **2460**.

Watson, I.: 2004, Minimum wages and employment: Comment, *Australian Economic Review* **37**(2), 166 – 172.

Weinzierl, M.: 2006, Estimating a relative utility function, *Working Paper* .

Wood, J. V.: 1996, What is social comparison and how should we study it?, *Personality and Social Psychology Bulletin* **22**(5), 520 – 537.

Zimmerman, D. W.: 2004, A note on preliminary tests of equality of variances, *British Journal of Mathematical and Statistical Psychology* **57**, 173 – 81.

Zizzo, D., Oswald, J. and Oswald, A.: 2001, Are people willing to pay to reduce others' incomes?, *Annales D'économie et de statistique* **63-64**, 39 – 65.

A Experimental instructions

Since we performed the experiments either at the Universität Karlsruhe (TU), Karlsruhe, Germany, or at the Max Planck Institute of Economics (MPI), Jena, Germany, the original instructions were written in German. In the following, we thus give translated instructions for each of the three experiments we described above. Additionally, formatting is changed to save some space, in one case a summarizing table the participants could use is left out, and instructions are given for only one treatment and participant (differences to other participants and treatments are minimal, of course). Instructions in the original formatting and language or for other participants and treatments can be obtained on request.

A.1 Spillover effects of minimum wages

Given below are instructions for employers in treatment MW.

Instructions:

You are participating in an experiment consisting of two parts. In the following these two parts will be named "part 1" and "part 2".

In this experiment, you can earn money, which will be paid out in cash immediately after the experiment. Your earnings depend on your own decisions and those of other participants. Every participant makes his decisions on his own at his computer box. Your anonymity will be guaranteed even after the end of the experiment. Communication between participants is not allowed. Please turn off your mobile phone and read the following instructions carefully. Please stay silently at your seat at the end of the experiment. We will call you individually and anonymously with the help of your box number and pay you off. We will have to exclude you from the experiment and all payments if you violate these rules.

In both parts of the experiment you form a group of four with three other participants. The composition of the group stays the same in both parts of the experiment. This means that in part 2 of the experiment you form a group with the same participants as in part 1.

Earnings are calculated in GE (Experimental Currency Unit) during the experiment. At the end of both parts of the experiments, the total earnings are calculated and converted into Euro at a fixed exchange rate. **This exchange rate is 100 GE = 80 Euro-Cent.**

Additionally to the earnings of the two parts of the experiments, participants once receive a fixed participation fee that does not depend on decisions. **This participation fee is: 500 GE.**

You are now handed out the instructions for the first part of the experiment. You will receive the instructions for the second part, after the first part is finished.

Part 1:

You form a group of four with 3 other participants. Three of you are employees (AN_1, AN_2, AN_3), one is the employer (AG). The role assignment was carried out randomly in the beginning of the experiment by the drawing of the box numbers. **Your role is that of an employer**.

In the following, employers and the employees interact for 5 periods. In each of these 5 periods, the employer offers wages to the employees who individually decide whether they want to be employed or not. If an employee is hired, he will produce a good for the employer. Before explaining the procedure of each of the 5 periods in detail, it is necessary to give the following information regarding the possibly produced goods. If, later on, AN_1 is hired by his employer, he will – automatically and without further decisions – produce a good which is immediately sold by the employer. The fixed selling price of this good is $R_1 = 100$ GE. Similar rules apply for AN_2. If he is hired, he will also automatically produce a good which is sold directly. The selling price of this good is also fixed and is $R_2 = 200$ GE. Analogically, the product possibly produced by AN_3 is sold for $R_3 = 300$ GE. **Each** of the 5 periods mentioned before is subdivided into **three stages** in the following way:

- Stage I: **Wage offers**
  The employer AG offers a wage to each employee AN, for which he (AG) is

willing to employ him (AN). The wage offer to AN$_i$ is denoted by w_i (with $i = 1, 2, 3$).
The employer is allowed to offer different wages, i.e., that, for example, the wage offer to AN$_1$ (w_1) does not have to be be identical to the wage offer to AN$_3$ (w_3).
However, the wage offers are not allowed to be lower than 1 and are also not allowed to be higher than the value of the good AN_i produces. Furthermore, only integers are allowed as wage offers. Overall, each of the three wage offers has to fulfill: $1 \leq w_i \leq R_i$ with w_i being an integer.
Altogether a total wage offer profile, (w_1, w_2, w_3), results that describes a wage offer to each of the three employees.

- Stage II: **Employment decision**
  Based on (w_1, w_2, w_3) each worker AN$_i$ is told only the wage offers to the both other workers, i.e, he does **not** know his own wage offer at first. Then, he names the limit t_i for his wage w_i above (or equal to) which he is willing to be employed. It has to fulfill: $t_i \leq R_i$.
  An example for AN$_2$: He learns w_1 and w_3 and then chooses his wage limit t_2.
  After this he learns his actual wage offer w_2. If his limit t_2 is less or equal to the wage offer w_2, he will be hired by the employer (for the wage w_2, see stage III). If his limit t_2 is higher than the wage offer w_2, he will not be hired in this period.

- Stage III: **Production, selling and earnings**
  If AN$_i$ is not being hired, nothing will be produced and sold and thus he and the employer will earn 0 GE from this potential hiring relationship. If AN$_i$ is hired, the good will be produced and sold and AN$_i$ will earn the offered wage w_i. The employer will earn the difference between the sales price and the paid wage, $R_i - w_i$. The employer's total earnings are the sum of the earnings from the hiring relationships with all workers. An example: If AN$_1$ and AN$_2$ are hired, but AN$_3$ is not, then AN$_1$ will earn w_1, AN$_2$ will earn w_2 and AN$_3$ will earn 0. The total earnings for the employer from the hiring relationships with the three workers follow as: $(R_1 - w_1) + (R_2 - w_2) + 0$.

In each period, all employees decide unaware of the decisions of the other employees. This means that, e.g., AN_1 does not know if AN_2 is hired or not before his own decision. However, each employee gets to know which employees had been hired und which not in the later course of each period.

After these three stages I to III the first periods ends. Then, another 4 periods are taking place with the same rules. The earnings of the 5 periods are summed up. After these 5 periods the second part of the experiment follows. You will receive new instructions for this part. Before we start the experiment, you have to answer some control questions at your computer box.

Part 2:

Just like in the first part of the experiment you form the same group of four with the same participants. Your role has also not changed. In the second part, another 5 periods take place that again consist of three stages I to III.

Before the first of the following 5 periods, additionally, a lower bound m will be set once and for all. This lower bound is set for each of the 5 following periods. The exact value of m will be given on your computer screen. This lower bound than applies to all wages, i.e., the employer AG has to offer a wage of at least m to each worker AN_i. Overall, each of the three wage offers now has to fulfill: $m \leq w_i \leq R_i$. Besides this the same rules for wage offers, employment decisions and earnings as in part 1 apply. The employees, however, have to name a limit at least as high as the lower bound m now, i.e.: $t_i \geq m$. Your total earnings are the sum of your earnings in parts 1 and 2 of the experiment plus the participation fee.

A.2 Employment protection and bullying

Given below are the instructions for employers in the treatment with weak employment protection and sanctions.

Instructions:

General rules:

You are participating in a decision experiment in which you can earn money. Please stop communicating with other participants from now on and turn off your mobile phone. Read the following instructions carefully. If you have a question, please raise your hand, and we will answer your question at your computer box. We will have to exclude you from the experiment and all payments if you violate these rules.

The instructions are identical for all participants except for the subsequent role assignment. Your anonymity will be guaranteed. This means that no other participant is going to learn your identity during or after the experiment.

In this experiment, **12 employees** and **8 employers** interact. The role assignment was carried out randomly in the beginning of the experiment. Every participant keeps his role during the whole experiment. **You are an employer.**
The whole experiment lasts for **10 periods**.

The earnings of each participant depend on his own decisions and those of other participants and are calculated in GE (Money Units) during the experiment. The employees once receive a participation fee of 50 GE at the beginning of the experiment. The employers receive an endowment in each period (see the following paragraphs). At the end of both parts of the experiments, the total earnings are calculated and converted into Euro at a fixed exchange rate. **This exchange rate is 1 GE = € .1.**

The first period:

The timing of the first period can be divided into 6 stages – stage I, II, III, IV, V and VI – as follows:

I: At the beginning of the first period, every employer receives a **basic endowment (*GR*) of 12 GE**. Then, every employer individually decides, whether he wants to hire an employee. This means that every employer can either hire 0 or 1 employee(s). Without knowledge of the employers' decisions, every employee individually decides, whether he wants to be hired or not. The employers who want to hire and the employees who want to be hired are then matched randomly.
If there are more employees who want to be hired than vacancies, some employees will not get a job. If there are more vacancies than employees who want to be hired, all employees will get a job, but some employers will not get an employee.

A hired employee produces a specific good for the employer. The **fixed wage** **(F)** the employer has to pay is set to **12 GE**. All employees who either were not hired or did not want to be employed receive a **payment** **(Z) of 7 GE**. For them the ongoing period ends after this payment.
All employers without an employee do not receive revenues from production. Thus, these employers only receive their basic endowment in this ongoing period.

The following stages (stages II -VI) apply to each pair of employer and hired employee.

II: The hired employee chooses his **effort level (e)** for the production of the good. Depending on this choice **costs** $K(e)$ arise. Overall, there are five effort levels. The respective costs of effort are given in Table A.1.

e	$K(e)$
$e1$	.0 GE
$e2$	.25 GE
$e3$	.75 GE
$e4$	1.5 GE
$e5$	2.5 GE

Table A.1: Costs of effort

III: The effort level influences the **output quantity Q** of the good and, thereby, the employer's profit. Overall **2, 3, 4 or 5** units of the good can be produced. The output quantity also depends on chance. The probability for a larger quantity increases with rising effort levels. The relationship between effort level and produced quantity is given in Table A.2.
For the first row, i.e., for effort level e1, the table can be explained as follows: If e1 is chosen, the quantity Q=2 is going to be realized with a probability of 50 percent, the quantity Q=3 with a probability of 30 percent, the quantity Q=4 with a probability of 15 percent and the quantity Q=5 with a probability of 5 percent. The other rows follow analogously.

In each period the randomness of the production affects all workers similarly, i.e, the same level of effort leads to the same output quantity, which is, however, determined by chance through the mechanism described above. An example: If a worker who has chosen e1 in the first period, produces three units of the good, then every other worker who chooses effort level e1, too, also produces 3 units.

Output quantity Q

	2	3	4	5
$e1$	.50	.30	.15	.05
$e2$	.40	.30	.20	.10
$e3$	.25	.25	.25	.25
$e4$	.10	.20	.30	.40
$e5$	.05	.15	.30	.50

Table A.2: Probabilities of output quantities for chosen levels of effort

IV: The employer learns which quantity of the good is produced in his firm, but not which effort level his employee has chosen. Additionally, the **average output quantity** **($\overline{Q}$)** over all matches of employer and hired employee is told to every employer (including those who do not hire an employee in this period). Then, every employer sells his produced goods at a **price** **(p)** **of 5 GE a piece**.

V: After learning the output quantity of his employee and the average output quantity, the employer can invest **up to 2 GE into punishing** his employee. He can choose every **punishment level** S between **0 GE** (no punishment) and **2 GE** (highest possible punishment) with the restriction that only one position after the decimal point is allowed. One unit of S decreases the employee's earnings threefold. For example, a level of punishment of 0.4 GE induces costs of 0.4 GE for the employer and costs of 1.2 GE for the employee.
In summary, the employer's profit G and the earnings A of an engaged employee follow as:

Employer: $G = GR - F + p \cdot Q - S$, so

$$G = 12 - 12 + 5 \cdot Q - S \text{ GE} .$$

Employee: $A = F - K(e) - 3 \cdot S$, so

$$A = 12 - K(e) - 3 \cdot S \text{ GE} .$$

VI: Now the employer decides if he wants to re-hire the employee. If he fires the employee, the next period will start in stage I for both of them again, i.e., with the decisions about hiring and working, respectively.
If he wants to re-hire the employee, the employee will have to decide whether he wants to continue to work for this employer. If he does not want to be re-hired, the next period will start in Stage I again for both of them.
The random matching algorithm in stage I automatically prohibits that employers and employees, who already formed a match in the previous period, are matched again.
If neither the employer nor the employee has ended the employment contract, the next period will start in stage II for both of them, i.e., with the effort choice by the employee.

Periods 2 to 10:

Basically, periods 2 - 10 are run like period 1. The only exception arises in stage VI. Here, the employers who have hired the same employee for at least two consecutive periods cannot fire this employee anymore. This means that an employer can fire his employee only in the first period of a particular match. From the second period of a match on this is not possible anymore. However, the employees are still allowed to quit their job. The last period automatically ends after stage V.

Final remarks:

The earnings of all 10 periods are summed up (and added to the participation fee for the employees). The earned GE will then be converted into Euro with the fixed exchange of € .1 per GE. The amount of Euro calculated this way is paid off individually and anonymously in cash.

Before we start the experiment, you have to answer some control questions at your computer box. If you are not able to answer a question, please lift your hand. We will then answer your question at your computer box.

Comment: We then additionally provided a list of all variables that is left out here due to space requirements.

A.3 Downsizing the labor force

Given below are the instructions for the L-employee, treatment AH.

Instructions:

■ **Experiment 1:**

1. General instructions:

Please stop communicating with other participants from now on and turn off your mobile phone. Read the following instructions carefully. If you have a question, please raise your hand, and the supervisors will answer your question at your computer box. We will have to exclude you from the experiment and all payments if you violate these rules. The instructions are identical for all participants except for the subsequent role assignment. Your anonymity will be guaranteed. This means that no other participant is going to learn your identity during or after the experiment. To begin with, you are taking part in an experiment consisting of **4 periods**. After this you will be given new instructions for another experiment!

In the first experiment, three participants will interact. Two of them will take the roles of employees, one will take the role of an employer. One of the employees is of type H (**H-employee**), the other of type L (**L-employee**). There is only one type of **employer**. The role assignment is carried out randomly in the beginning of the experiment. Each participant keeps his role during the whole experiment. **You are an L-employee.**

The earnings of every participant depend on his or her own decisions and those of the other participants. Earnings are calculated in ECU (Experimental Currency Unit) during the experiment. At the end of both experiments, they will be converted into Euro at a fixed exchange rate. **This exchange rate is 30 ECU = 1 Euro.**

Additionally, participants receive a fixed participation fee that does not depend on decisions, but will be offset against payoffs if necessary. **This participation fee is: 90 ECU.**

2. Periods 1 and 2:

2.1 General rules:

Each triplet of participants, consisting of an employer, an H-employee, and an L-employee, interacts for 4 periods. Each of the two periods 1 and 2 is basically constructed as follows:

1. The employer offers one contract that applies to both employees. It consists of two components: a **fixed wage** W with $24 \leq W \leq 40$ and a **piece rate** r (with $0 \leq r \leq 20$) that must be paid for each unit of output. Furthermore, the employer has to make conjectures about the employees' effort levels (see 2.). Up to 1 decimal place is allowed for each of the inputs named above.
2. Knowing the offered contract, each employee independently chooses an effort level, i.e., the H-employee chooses e_H, the L-employee chooses e_L. Restrictions are: $0 \leq e_H \leq 10$ and $0 \leq e_L \leq 10$. Again, up to 1 decimal place is allowed. One unit of effort leads to exactly one unit of output the employer is selling. The **gross output** Q thus equals the sum of chosen efforts.

This ends the interactions of a period. Payoffs result as follows:

- Employer: $(24 - r) \cdot (e_H + e_L) - 2 \cdot W$.
- H-employee: $W + r \cdot e_H - 1 \cdot (e_H)^2$.
- L-employee: $W + r \cdot e_L - 6 \cdot (e_L)^2$.

After each period every participant gets to know effort levels, gross output, and payoffs of all participants.

2.2 Calculator:

Additionally, the software provides a **calculator** to each participant. You can use this calculator for two minutes in every period, after which you have to make your decision at the latest. The calculator allows every participant

- to calculate the employer's payoff for various levels of W and r and various effort levels e_H and e_L, and
- to calculate the employees' payoffs for various levels of W and r and various effort levels e_H and e_L.

Note that as an employer you are only able to make conjectures about effort levels since you do not know the employees' decisions yet. As an employee, however, you know the decisions of the employer. They are preset in the calculator. You will not learn the employer's conjectures.

2.3 Additional restrictions:

As an **additional restriction for the employees**, you are limited to choose effort levels – e_H or e_L – that guarantee that payoffs are larger than or equal to zero in each period. You can check this restriction with the help of the calculator.

As an **additional restriction for the employer**, you are limited to offers W and r that, in addition to the conjectures about effort levels e_H and e_L, also given by yourself, guarantee that expected payoffs are larger than or equal to zero in each period. You can check this restriction with the help of the calculator.

These restrictions imply that period payoffs smaller than zero are only possible for employers, e.g. if effort levels are below conjectured efforts. However, the employer is able to restrict this risk by choosing W and r appropriately; payoffs larger zero should be the norm. Payoffs of employers and employees are summed up over the first two periods that are played as described above.

3. Periods 3 and 4:

Before the third period, each employee may choose between two alternatives:

I. to **keep the L-employee** or

II. to **lay off the L-employee**.

The H-employee will always be kept.

If the employer hires the L-employee again (case I), periods 3 and 4 are played analogously to periods 1 and 2.

If the employer lays off the L-employee (case II), the L-employee will receive a payment of 15 ECU from the experimenters (not from the employer) in each of the periods 3 and 4. Consequently, the L-employee does not make any decisions and does not learn the other participants' payoffs in periods 3 and 4. In each of the periods of this case (II), the employer offers a new contract to the H-employee only. The same bounds for contracts and effort levels apply. The payoff of the H-employee is calculated as before. Of course, the employer now earns $(24-r)\cdot e_H - W$. Restrictions are unchanged; calculators are provided again.

Payoffs of periods 3 and 4 are added to those of periods 1 and 2 and to the participation fee, are converted into Euro, and are paid out anonymously and in cash at the end of both experiments. If the employer's payoff from periods 1 to 4 is smaller than zero, it will be subtracted from the participation fee. If the rest is smaller than zero, it will be offset against the payoffs from the other experiment. If there is still a debt, this has to be paid for at the end of both experiments – either in cash or by administrative work. Please note again as an employer that this situation can be avoided almost completely by choosing W and r appropriately; payoffs larger zero should be the norm. The employees' payoffs are always larger than or equal to zero.

In the following, last experiment, you will not interact with the same participants as in this experiment again. Before we start the experiment, you have to answer some control questions.

■ **Experiment 2:**

We will now repeat the same experiment one more time, i.e., **all 4 periods** are played again. This means that, again, you will receive your **participation fee** and additional payments, depending on your decisions. Payoffs of all periods are added, converted, and paid out as before. Furthermore, you keep the same role as in the previous experiment, but it is guaranteed that no one will be matched with the same participants again.

Please stay silently at your seat at the end of the experiment until we call you individually and anonymously with the help of your box number and pay you off.

B Theoretical addenda – Spillover effects of minimum wages

B.1 Wage ordering

We want to discuss here under which circumstances the intuitive wage profile **(A.6.1)** $w_L^* < w_M^* < w_H^*$ holds. This clearly is not possible without relying on more specific assumptions about reservation utilities and concrete utility functions.

We start by introducing a second *reservation utility assumption*: **(A.1.a)** $r_L < r_M < r_H$. It states that the reservation utilities' ranking mirrors the ranking of productivity differences. We think this assumption is less strict than it might look at first glance, since it specifically does not require reservation utilities to be less divergent than marginal revenues. It does, however, require comparability of agents' utility functions. Alternatively, one can re-interpret the reservation utilities as pie shares participants commonly demand in ultimatum games adjusted for relative income preferences. Of course, this demands similar utility functions like ACU and RCU where utility directly depends on own and others' wages (see the following paragraphs). FFZ2006 use the heuristic of about 30 to 40 percent of the total pie size to predict firms' offers. Offers below this are most often rejected. Since pies are unequal in our game, one could generalize this by expecting reservation utilities to be larger for players who bargain over larger pies.

However, **(A.1.a)** alone does not fully determine the wage profile. We also need some kind of similarity in preferences which will be clearer with an example. For expositional purposes, assume only a low and a medium productive player interact $(i, j \in \{L, M\};\ i \neq j)$. Their RCU functions shall be given by

$$u_L \;=\; w_L^{1-\alpha_L} \cdot \left(\frac{w_L}{w_M}\right)^{\alpha_L} \tag{B.1}$$

and

$$u_M \;=\; w_M^{1-\alpha_M} \cdot \left(\frac{w_M}{w_L}\right)^{\alpha_M} \tag{B.2}$$

We first equate u_L to r_L and u_M to $r_M = r_L + d$ (with $d > 0$) since this should hold in equilibrium. Now suppose we want to check whether the principal can set the wages to $w_L = w_M$. The second factors of u_L in (B.1) and u_M in (B.2) then vanish. We can now solve both equations for r_L and equate them. Demanding equal wages for M and L then requires

$$w_L^{1-\alpha_L} = w_L^{1-\alpha_M} - d\ . \tag{B.3}$$

With similar RCU functions Johansson-Stenman et al. (2002) estimated α for different income groups. They did not find significant differences in mean values of α with α ranging from .31 to .43. Though this may be different on individual level and the estimations were not explicitly based on different skill levels, we use this best heuristic available to us and assume α to be quite similar for L and M. Then, it is immediately obvious that (B.3) can never hold for all d significantly larger than zero. This implies that the principal cannot set w_L equal to w_M. Furthermore, $w_L > w_M$ is also impossible, since this would further increase the left, and further decrease the right side of equation (B.3).

An even stronger support for the result that $w_L = w_M$ cannot hold, is given by the ACU function parameter estimates reported by Johansson-Stenman et al. (2002). The authors found that α was increasing for higher income levels, which, if transferred to an ACU analogon of (B.3) leads to a further increase of the right side of the equation.

The reader might also argue that the terms in brackets on the right sides of (B.1) and (B.2) should be modified by a more sophisticated reference standard than equal wages. With x_L and x_M as such converters, (B.3) becomes

$$w_L^{1-\alpha_L} x_L^{\alpha_L} = w_L^{1-\alpha_M} x_M^{\alpha_M} - d\ . \tag{B.4}$$

With our productivity differences at least $0 < x_M < 1 < x_L$ should hold, what further increases the left and decreases the right side (although one might now consequentially expect d to decrease or even vanish).

To sum things up: With this expanded set of assumptions, concrete RCU (or ACU) functions suggest that the equilibrium wage profile likely mirrors the productivity differences, i.e., $w_L^* < w_M^* < w_H^*$ holds.

B.2 Comparative statics

The first-order conditions (FOC) of the *Lagrangian function* $L(w_M, w_H, \lambda_M, \lambda_H; m)$ are denoted as follows: FOC no. 1: $\partial L/\partial w_M = 0$; no. 2: $\partial L/\partial w_H = 0$; no. 3: $\partial L/\partial \lambda_M = 0$; no. 4: $\partial L/\partial \lambda_H = 0$.

They implicitly define four equations $F^p(w_M,\ w_H,\ \lambda_1,\ \lambda_2;\ m)$ with p being the same number as in the corresponding FOC. This, in turn, describes all four endogenous variables depending on the exogenous minimum wage m only: $w_M^{**} = f^1(m)$, $w_H^{**} = f^2(m)$, $\lambda_M^{**} = f^3(m)$, $\lambda_H^{**} = f^4(m)$. Differentiating again regrouping, and some defining then leads to:

$$\begin{bmatrix} \frac{\partial F^1}{\partial w_M} := A & \frac{\partial F^1}{\partial w_H} := B & \frac{\partial F^1}{\partial \lambda_M} := C & \frac{\partial F^1}{\partial \lambda_H} := D \\ \frac{\partial F^2}{\partial w_M} := E & \frac{\partial F^2}{\partial w_H} := F & \frac{\partial F^2}{\partial \lambda_M} := G & \frac{\partial F^2}{\partial \lambda_H} := H \\ \frac{\partial F^3}{\partial w_M} = C & \frac{\partial F^3}{\partial w_H} = G & \frac{\partial F^3}{\partial \lambda_M} = 0 & \frac{\partial F^3}{\partial \lambda_H} = 0 \\ \frac{\partial F^4}{\partial w_M} = D & \frac{\partial F^4}{\partial w_H} = H & \frac{\partial F^4}{\partial \lambda_M} = 0 & \frac{\partial F^4}{\partial \lambda_H} = 0 \end{bmatrix} \begin{bmatrix} \frac{dw_M}{dm} \\ \frac{dw_H}{dm} \\ \frac{d\lambda_M}{dm} \\ \frac{d\lambda_H}{dm} \end{bmatrix} = \begin{bmatrix} -\frac{\partial F^1}{\partial m} := I \\ -\frac{\partial F^2}{\partial m} := J \\ -\frac{\partial F^3}{\partial m} := K \\ -\frac{\partial F^4}{\partial m} := L \end{bmatrix}$$

or in simpler vector form: $M \cdot N = O$.

With Cramer's Rule, the desired comparative static results are

$$\frac{dw_M^{**}}{dm} = \frac{|M_1|}{|M|}$$

and

$$\frac{dw_H^{**}}{dm} = \frac{|M_2|}{|M|} ,$$

where $|M_i|$ denotes the determinant of vector M with the i-th column replaced by vector O.

Due to the fact that the four lower right cell entries all equal zero, all second-order derivatives (in A, B, E, and F), vanish and some further simplifications yield

$$\frac{dw_M^{**}}{dm} = \frac{G \cdot L - H \cdot K}{D \cdot G - C \cdot H} \qquad \text{and} \qquad \frac{dw_H^{**}}{dm} = \frac{D \cdot K - L \cdot C}{D \cdot G - C \cdot H} ,$$

which directly gives (2.5) and (2.6). *q.e.d.*

B.3 Total differential, utility of worker L

For the RCU functions defined by (2.8) the two comparative static derivatives (after some cancellations) are:

$$\frac{dw_M^{**}}{dm} = \frac{\beta_M \beta_H + \alpha_M}{1 - \beta_M \alpha_H} \cdot \frac{w_M}{w_L} \quad \text{and} \quad \frac{dw_H^{**}}{dm} = \frac{\alpha_M \alpha_H + \beta_H}{1 - \beta_M \alpha_H} \cdot \frac{w_H}{w_L} .$$

Substituting this and all other terms into (2.7) and further cancellations lead to:

$$\frac{du_L}{dm} \geq 0 \Leftrightarrow \frac{1 - \alpha_L(\alpha_M + \beta_M \beta_H) - \beta_L(\alpha_M \alpha_H + \beta_H) - \alpha_H \beta_M}{1 - \beta_M \alpha_H} := \frac{N}{D} \geq 0 . \quad \text{(B.5)}$$

Repeating the same steps for the ACU functions defined by (2.9) leads to exactly the same intermediate result. Since the denominator of (B.5) is larger than zero for either $\beta_M \neq 1$ or $\alpha_H \neq 1$ (which we implicitly assume to avoid further complexities), the numerator N determines the sign of (B.5). Substituting $\beta_L = 1 - \alpha_L$, $\beta_M =$

$1-\alpha_M$, $\beta_H = 1-\alpha_H$ gives:

$$N = 1-(1-\alpha_M)\alpha_H-\alpha_L\alpha_M-\alpha_L(1-\alpha_M)(1-\alpha_H)-(1-\alpha_L)(1-\alpha_H)-(1-\alpha_L)\alpha_M\alpha_H\ ,$$

which is equal to 0.

Furthermore, this equation immediately tells us that $N > 0$ holds if any $1-\alpha_i-\beta_i$ is larger than 0. *q.e.d.*

The approach to directly investigate the sign of (2.7) does not lead to an unifying categorization for RCU and ACU functions. Simple calculations only yield that as long as the derivative $\partial u_i/\partial w_i$ is larger than the absolute value of $\partial u_i/\partial w_j + \partial u_i/\partial w_k$, the total differential is larger than zero, but this is only a sufficient condition that solely the ACU necessarily functions fulfill (with $\partial u_i/\partial w_i = 1$ and $\partial u_i/\partial w_j+\partial u_i/\partial w_k < 1$). But even adding factors $x_{i,j}$ into the brackets in (2.8) does not change results for RCU functions. For example, the converters $x_{i,j} = R_j/R_i$ would constitute productivity differences as agents' new aspirations level. However, every set of converters (not only the example named above) leads to (B.5) again.

C Statistical addenda – Downsizing the labor force

C.1 List of tests for differences between treatments AH and UH

All test results for contract offers and payoffs are given. All tests are performed as independent two sample test. Sample sizes are 48, 35, and 13 for all firms, only firms D, and only firms ND in treatment AH and 16, 13, and 3 for treatment UH. In Tables C.1, C.2, and C.3 below, we abbreviate the performed tests as follows: Mann-Whitney U-tests (U), U-tests and robust rank-order tests (U/rro).

	a) All firms	
	Before	After
F	.102 (.115) (U/rro)	.486 (U)
r	.577 (.565) (U/rro)	.510 (U)
Π	.997 (U)	.444 (U)
ω_h	.664 (U)	.526 (U)
ω_l	.307 (U)	.489 (U)

Table C.1: Treatment UH, all firms, main results

	b) Only firms D	
	Before	After
F	.223 (.255) (U/rro)	.681 (U)
r	.950 (U)	.209 (.192) (U/rro)
Π	.555 (.567) (U/rro)	.320 (.292) (U/rro)
ω_h	.977 (U)	.238 (.219) (U/rro)
ω_l	.385 (U)	

Table C.2: Treatment UH, only firms D, main results

	c) Only firms ND	
	Before	After
F	.489 (U)	.500 (U)
r	.680 (.709) (U/rro)	.536 (U)
Π	.818 (.814) (U/rro)	1.000 (U)
ω_h	.704 (.710) (U/rro)	.439 (U)
ω_l	.637 (U)	.800 (U)

Table C.3: Treatment UH, only firms ND, main results

C.2 Logistic regression results, treatments AH and AL

Given below are translated, relevant output tables for the logistic regression of the dependent dummy variable *Down* defined in subsection 4.5.2 for treatment AH (Tables C.4 - C.7) and treatment AL (Table C.8). The tests were performed using the statistic software SPSS 16. Due to the few undefined values of $reldev_h$, the final regression for treatment AH was performed for only of 46 of 48 values. The results differ negligibly when using all 48 observations or only Π.

We used a forward algorithm to check whether to include any of the 17 explanatory variables we observed before downsizing. These variables were the fixed wage, the piece rate, both efforts, the three payoffs, the absolute and relative effort deviations as well as the effort conjectures, and their absolute and relative deviations from optimum effort. This algorithm uses score tests to decide which variable to include next as well as likelihood ratio test to evaluate whether the inclusion improves the model's explanatory power significantly. The algorithm stops after step 1, including only the firm's payoff Π as explanatory variable. Table C.7 gives the results of the likelihood ratio omnibus test for this first step compared to the null model without explanatory variables. The inclusion of Π contributes significantly ($p = .002$). The effect of the principal's profit is significant (Wald test: $p = .017$, see Table C.4).

For step 0 of treatment AH, where predictions suggest that downsizing took place, 33 out of 46 observations are predicted correctly, a ratio of 71.7%. This classification table is omitted due to space restrictions. After step 1, the model predicts 78.3 % of observations correctly, which is an improvement over the null model.

Variables in the equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	Π	-.075	.032	5.666	1	.017	.928
	Constant	3.245	1.150	7.957	1	.005	25.663

Table C.4: Treatment AH, logistic regression: Relevant variables

Classification Table

			Predicted Down (1=yes)		Percentage of correct predictions
		Observed	.00	1.00	
Step 1	Down (1=yes)	.00	5	8	38.5
		1.00	2	31	93.9
		Overall			78.3
The cut value is .500					

Table C.5: Treatment AH, logistic regression: Classification Table

Variables not in the equation

			Value	df	Sig.
Step 1	Variables	F	.277	1	.599
		r	.076	1	.783
		$conj_h$	.000	1	.992
		$conj_l$	.071	1	.790
		e_h	.476	1	.490
		e_h	.396	1	.529
		ω_h	.027	1	.870
		ω_l	.435	1	.510
		$absdev_h$	2.769	1	.096
		$absdev_l$	.355	1	.551
		$absconj_h$	.672	1	.412
		$absconj_l$	.006	1	.938
		$reldev_h$	.549	1	.459
		$reldev_l$	.326	1	.568
		$relconj_h$	.019	1	.889
		$relconj_h$	.003	1	.955

Table C.6: Treatment AH, logistic regression: Other variables

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	10.021	1	.002
	Block	10.021	1	.002
	Model	10.021	1	.002

Table C.7: Treatment AH, logistic regression: Omnibus Tests

Variables not in the equation

			Value	df	Sig.
Step 0	Variables	F	.010	1	.919
		r	.167	1	.683
		$conj_h$	2.760	1	.097
		$conj_l$	.974	1	.324
		e_h	.615	1	.433
		e_l	.060	1	.806
		Π	1.706	1	.192
		ω_h	.101	1	.751
		ω_l	.170	1	.680
		$absdev_h$	.681	1	.409
		$absdev_l$	.257	1	.612
		$absconj_h$	1.408	1	.235
		$absconj_l$	.761	1	.383
		$reldev_h$	1.449	1	.229
		$reldev_l$	.441	1	.507
		$relconj_h$	2.906	1	.088
		$relconj_l$	.171	1	.680

Table C.8: Treatment AL, logistic regression

List of Figures

2.1 Treatment MW, average wage offers 40
2.2 Treatment MW, relative changes in paid wages 46
2.3 Treatment MW, relative changes in wage thresholds 49
2.4 Treatments MW and CTR, relative changes in paid wages and wage thresholds 55

3.1 Effort choices, matching periods 77
3.2 Effort choices, long- and short-term matches 78
3.3 Sanctions, matching periods 81
3.4 Sanctions, long- and short-term matches 82

List of Tables

2.1 Hiring frequencies and payoffs . . . 41
2.2 Treatment MW, paid wages and successful wage thresholds . . . 42
2.3 Treatment MW, all wage offers and wage thresholds . . . 47
2.4 Treatment MW, wage thresholds and revenue relations . . . 51
2.5 Treatment CTR, paid wages . . . 53
2.6 Treatment CTR, wage thresholds . . . 56

3.1 Treatment overview . . . 66
3.2 Probability functions X_1, X_2, X_3, X_4, X_5 . . . 67
3.3 Match lengths, all treatments . . . 73
3.4 Average efforts and sanctions, matching periods . . . 75
3.5 Sanction frequencies . . . 79
3.6 Efforts, Outputs, Sanctions, and Earnings, with and without firing . 84
3.7 Sanction frequency, with and without firing . . . 85
3.8 Sanction severeness categories, with and without firing . . . 86
3.9 Frequencies of firing employers and quitting agents . . . 88
3.10 Average payoffs, matching periods . . . 90

4.1 All treatments, downsizing decisions . . . 104
4.2 Treatment AH, main results . . . 105
4.3 Treatment AH, effort deviations . . . 106
4.4 Treatment AL, main results . . . 109
4.5 Treatment AL, effort deviations . . . 110
4.6 Treatment UH, main results . . . 113
4.7 Treatment UH, effort deviations . . . 113

A.1 Costs of effort 138
A.2 Probabilities of output quantities for chosen levels of effort 139

C.1 Treatment UH, all firms, main results 150
C.2 Treatment UH, only firms D, main results 151
C.3 Treatment UH, only firms ND, main results 151
C.4 Treatment AH, logistic regression: Relevant variables 152
C.5 Treatment AH, logistic regression: Classification Table 152
C.6 Treatment AH, logistic regression: Other variables 153
C.7 Treatment AH, logistic regression: Omnibus Tests 153
C.8 Treatment AL, logistic regression 154

List of Abbreviations

Renaming all abbreviations here, especially those of all variables, would certainly be too much. Instead we list the most important abbreviations again and order them by topic.

Experimental treatments:

AH	announced downsizing, high incentive-treatment (Chapter 4)
AL	announced downsizing, low incentive-treatment (Chapter 4)
CTR	control treatment (Chapter 2)
MW	minimum wage treatment (Chapter 2)
tI	treatment I: no employment protection and sanctions (Chapter 3)
tII	treatment II: weak employment protection and sanctions (Chapter 3)
tIII	treatment III: strict employment protection and sanctions (Chapter 3)
tIV	treatment IV: no employment protection and no sanctions (Chapter 3)
tV	treatment V: weak employment protection and no sanctions (Chapter 3)
tVI	treatment VI: strict employment protection and no sanctions (Chapter 3)
UH	unannounced downsizing, high incentive-treatment (Chapter 4)

Literature:

AHW2003	Andreoni et al. (2003)
FFZ2006	Falk et al. (2006)
FHM2008	Falk et al. (2008)

Statistics:

mp-rule	each matching period is used as one observation
one-match-rule	each match is used as one observation
rro test	robust rank-order test
U-test	Wilcoxon Mann-Whitney U-test
Wilcoxon test	Wilcoxon paired sample test
WS test	Welch-Satterthwaite test

Models, Miscellaneous:

ACU	Additive Comparison Utility
firms D	firms that downsize
firms ND	firms that do not downsize
EPL	employment protection laws/legislation
RCU	Ratio Comparison Utility