

Monitoring and Pay: An Experiment on Employee Performance under Endogenous Supervision^a

Dennis A.V. Dittrich^b Martin G. Kocher^c

This version: 26 September 2006

We present an experimental test of a shirking model where monitoring intensity is endogenous and effort a continuous variable. Wage level, monitoring intensity and consequently the desired enforceable effort level are jointly determined by the maximization problem of the firm. As a result, monitoring and pay should be complements. In our experiment, between and within treatment variation is qualitatively in line with the normative predictions of the model under selfishness assumptions. Yet, we also find evidence for reciprocal behavior. The data analysis shows, however, that it does not pay for the employer to rely on the reciprocity of employees.

Keywords: incentive contracts, supervision, efficiency wages, experiment, incomplete contracts, reciprocity

JEL-Classification: C91, J31, J41

^a Financial support by the Center for Experimental Economics at the University of Innsbruck (sponsored by Raiffeisen Landesbank Tirol AG) is gratefully acknowledged.

^b University of Erfurt. Address: Nordhäuser Str. 63, D-99089 Erfurt, Germany. email: dennis.dittrich@uni-erfurt.de

^c University of Amsterdam and University of Innsbruck. Address: Roetersstraat 11, NL-1018 WB Amsterdam, Netherlands. email: martin.kocher@uibk.ac.at

1 Introduction

While incomplete contracts between principals and agents have been studied extensively in economics both from a theoretical and an empirical viewpoint, the important connection between payments and monitoring in such relationships has attracted by far less attention. Given the obvious practical importance of an optimal mix of direct monetary incentives (like wages in an employer-employee relation) and monitoring, the scarcity of empirical work is intriguing. It is even more astonishing in light of the fact that the intuitive notion according to which monitoring and pay are substitutes is far less obvious than it seems at first sight. Conventional wisdom would predict that workers who cannot be monitored properly have to be well paid in order to avoid shirking, and conversely, workers who act under close scrutiny do not have to be paid above market-clearing levels. However, whether wages and supervisors are substitutes or complements has long been an unclear issue both theoretically and empirically, although it is obviously key for understanding labor markets with imperfectly enforceable contracts.¹

In this paper we will rely on laboratory experiments to empirically assess incomplete contracts that incorporate a possibility for costly monitoring. On a posted offer market principals can offer contracts that specify a wage and a monitoring probability following a shirking detection technology as well as a desired non-binding effort level to be exerted by the agent. After having accepted a contract, agents have to submit an effort level, and a random mechanism according to the monitoring probability determines whether the agent is actually monitored. We implement a simple static model of the shirking version of the efficiency wage hypothesis in which the short side of the market is labor demand. As already mentioned, our main focus is on the important question whether wages and supervisors are indeed substitutes as intuition predicts or rather complements. For this end, four treatments with different parameterizations of productivity and monitoring costs are put to an experimental test. These four settings can be viewed as representing four different industries or economic branches that naturally exhibit different combinations of productivity and monitoring costs.

The experimental design that we use is related to Fehr, Kirchsteiger and Riedl (1996), Fehr, Gächter and Kirchsteiger (1997), Fehr, Klein and Schmidt (2001), and Fehr and Gächter (2002). Two distinctive features of our approach render our results, however, much more general: (i) we implement, as far as we know for the first time, continuous effort (costs) for the agent,

¹ The terms "principal" and "employer" as well as "agent" and "employee" are used interchangeably throughout the paper. The same holds for the expressions "monitoring" and "supervision" as well as "pay" and "wage".

and (ii) we apply an endogenous monitoring technology. Both features are desirable from a theoretical viewpoint as well as from a more applied perspective, for reasons that will be expounded in detail below.

On the theoretical side, it was Allgulin and Ellingsen (2002), who first challenged the conventional wisdom of the substitutive relation between wages and monitoring. Their model shows that monitoring and the level of pay are negatively related, i.e. substitutes, only under very restrictive assumptions that have, however, been common in the previous literature: (i) workers only have a choice between two effort levels (working or shirking), and (ii) the desirable level of a worker's effort is given exogenously. In the more general case, in which workers can choose from a continuum of effort levels and in which the desirable level of effort emerges from the solution to the firm's profit maximization, complementarity of supervision and pay results, regardless of which of the model's parameters is varied.

However, to what extent these specific theoretical results are indicative for actual behavior of subjects in such principal-agent relations is an open question. Putting the issue of some critical assumptions in the prevalent models aside for a moment, in reality, behavioral regularities like reciprocity (a positive wage- or rent-effort correlation) or, more general, social preferences have been shown to play an important role in shaping principal-agent relationships in a lot of existing experiments (among the first Fehr, Kirchsteiger and Riedl, 1993). The existence of social preferences might completely turn the picture with regard to the complementarity of monitoring or pay. A sufficient degree of reciprocity, for instance, may render any monitoring simply unnecessary.

Therefore, it is ultimately a matter of empirical analysis to substantiate whether monitoring and pay are actually substitutes or complements (see also, e. g., Chang and Lai, 1999; Demougin and Fluet, 2001). A natural approach to answer the question at hand would of course be the collection of field data. Unfortunately, by using field data one has to cope with quite a few serious problems that are common to tests of efficiency wage models. For instance, with field data the accuracy of monitoring is difficult to measure properly, effort levels are not easy to classify, and entry costs may create a severe sample selection bias (see Allgulin and Ellingsen, 2002). Therefore, also Prendergast (1999) argues that it is very difficult to assess the issue with data from the field (p. 45): "The problem is that either may easily arise in a world of efficiency wages and depends critically on the source of variation across firms. On the one hand, if the source of variation across firms is the cost of supervisors, then the two instruments are likely to be substitutes, where firms substitute away from high-cost supervisors into wages. On the other hand, if the source of variation across firms is in the return to effort (so some firms value effort exertion more than others), those firms that want more effort will use more of both in-

struments relative to those that do not values such high effort." Using an experiment is, thus, obviously helpful and at least a valuable complement to existing empirical studies that rely on field data.

The results of our experiment reveal that although wages are generally too high and the monitoring probability is too low in comparison to the theoretical predictions under selfishness assumptions, participants behavior is qualitatively in line with theory, especially when we look at the treatment effects. Analyzing the within treatment variation, we find a clear and highly significant positive correlation between monitoring and pay. Thus, the predictions of the shirking model are confirmed, and monitoring and supervision are indeed complements. Yet, we also observe clear evidence for the existence of reciprocity, for which the too high wages and too low monitoring intensity are first indications. But also more elaborate tests confirm the existence of reciprocity among principals and agents. It is, however, important to note that relying on reciprocity alone is not an efficient strategy for principals. The principals' earnings are at least as high when they design a contract with enforceable effort as when they invested the same amount of money in the labor relationship but offer a higher rent and monitor less. A striking feature of the data is that, although the reciprocity hypothesis is confirmed, contracts without monitoring fare particularly badly. A low intensity of monitoring, however, goes a long way. Another important finding in line with the existing literature is that the degree of reciprocity is highly heterogeneous across subjects. In other words, the rent-effort relation is by no means the same for all participants.

The remainder of the paper is laid out as follows. In section 2 we present briefly Allgulin and Ellingsen's (2002) shirking model with continuous effort and endogenous monitoring levels on which our experiment relies. Sections 3 reviews existing empirical and experimental evidence on our research question in turn. The details of our experimental design, our hypotheses and the laboratory protocol are presented in sections 4 and 5. Section 6 is devoted to the presentation of the results of our experiment, and section 7 discusses implications of them in the context of existing empirical and experimental studies.

2 The shirking model

The general shirking model of the efficiency-wage hypothesis is motivated by the dynamic dimension of the labor relationship. It assumes that the principal proposes a contract $\{w_t; t = 0, 1, \dots, +\infty\}$, specifying the wage the employee will receive at each date t . If the agent is caught shirking she is paid to the end of the period and fired, then. Note that the shirker receives her wages for that last period even if she has not supplied any effort. This assumption is

an offshot of the unverifiable character of production, which prevents employers from proposing a remuneration solely based on results. The optimal contract of this dynamic model implies zero rent for the worker at date $t = 0$. But in all subsequent periods, the agent obtains a utility that is strictly greater than her utility from being unemployed (see Cahuc and Zylberberg, 2003).

In the following we present a simplified static version of the shirking model in which optimal effort and supervision intensity arise endogenously. We consider a bilateral interaction between a risk-neutral principal and a risk-neutral agent.² Both maximize their own utility. The principal wants to delegate some work \tilde{e} to an agent and offers her a compensation $w(e)$. Further she can invest some money $\mu M(p)$ in a monitoring technology that allows her to verify with an investment-dependent probability p whether the agent falls short of the desired effort \tilde{e} . Effort determines the principal's benefit $\beta B(e)$ at some cost to the agent $\zeta C(e)$. We make the following standard assumptions regarding the functional forms: (i) $B'(e) > 0$, $B''(e) \leq 0$, (ii) $C(0) = 0$, $C'(e) > 0$, $C''(e) > 0$, (iii) $p'(s) > 0$.

The ex-post utility of the principal can be expressed by

$$\Pi = \beta B(e) - w(e) - \mu M(p), \quad (1)$$

and the ex-post utility of the agent is

$$U = w(e) - \zeta C(e). \quad (2)$$

The compensation $w(e)$ has a lower limit \underline{w} which may be due to a wealth constraint or legal rules. Further, since effort is not always observed, the compensation contract has to specify some payment \bar{w} that the agent receives in this case. Though it is irrelevant for the agent's incentives we assume $\bar{w} = w(\tilde{e})$ as is common in the literature. However, the principal would have an incentive not to monitor if we allowed $\bar{w} < w(\tilde{e})$. The agent maximizes expected utility,

$$E[U] = pw(e) + (1 - p)\bar{w} - \zeta C(e). \quad (3)$$

² Assuming risk-neutrality is also innocuous when deriving specific theoretical predictions for our experiment. If experimental participants were not close to risk-neutral over the monetary domain at stake in experiments, we would have to accept ridiculously high risk-aversion levels for higher stake levels according to calibration results in Rabin (2000).

The following incentive compatibility constraint must be satisfied for all e if the principal is free to induce any level of effort \tilde{e} :

$$pw(\tilde{e}) + (1 - p)\bar{w} - \zeta C(\tilde{e}) \geq pw(e) + (1 - p)\bar{w} - \zeta C(e). \quad (4)$$

A step function of the form $w(e) = \underline{w}$ for $e < \tilde{e}$ and $w(e) = \bar{w}$ for $e \geq \tilde{e}$ can replicate any incentive compatible contract that implements \tilde{e} without loss to the principal, i. e. the agent gets \bar{w} if she meets or exceeds the target and the minimum payment \underline{w} otherwise. If an agent wants to deviate, she will always deviate to $e = 0$. The incentive compatibility constraint thus becomes

$$p(\bar{w} - \underline{w}) \geq \zeta C(\tilde{e}). \quad (5)$$

We assume that an indifferent agent will exert the desired effort level. Thus, the incentive compatibility constraint becomes an equality from which we obtain the actual effort the principal will be able to enforce,

$$e(p, \bar{w}) = C^{-1}((\bar{w} - \underline{w})p/\zeta). \quad (6)$$

Her problem, then, is to find a probability p and a wage w to maximize

$$\Pi(p, w) = \beta B(e(p, w)) - w - \mu M(p) \quad (7)$$

subject to $w \geq \underline{w}$ and $p \in [0, 1]$. Let $r(e) = \beta B'(e)/\zeta C'(e)$, then the first order conditions for the solution are

$$p^* r(e^*) - 1 \leq 0 \quad \text{with equality if } w^* > \underline{w}, \text{ and} \quad (8)$$

$$(w^* - \underline{w})r(e^*) - \mu M'(p^*) \geq 0 \quad \text{with equality if } p^* < 1. \quad (9)$$

Allgulin and Ellingsen (2002) prove the following propositions. First, the marginal benefit from increased effort will be larger than the marginal cost whenever the principal chooses to monitor imperfectly. Second, there must be a positive level of monitoring in order to induce any effort. Third, sufficient conditions for monitoring and pay to be complementary instruments are: (i) the principal's benefit function $B(e)$ and the agent's cost of effort function $C(e)$ are both represented by power functions; (ii) the principal's benefit function $B(e)$ is linear, and the relative growth of costs of effort is decreasing in the effort level; (iii) if the source of variation is β or μ , monitoring and pay are complementary instruments if and only

if $-p^* M''(p^*)/M'(p^*) < 1$.

Let us also analyze the interaction in the presence of reciprocity and/or fairness. Although we do not intend to go into the details of various theories that take social preferences into account, it may be helpful to briefly provide an intuitive prediction – though based on a rigorous model – of the impact of reciprocity or fairness on behavior in the principal-agent relationship.

Intention-based fairness models (starting with Rabin, 1993) would predict that people who are motivated by reciprocal fairness are willing to sacrifice resources to be kind to others who are perceived to act kindly (positive reciprocity) and to be unkind to or to punish those who are perceived to act unkindly (negative reciprocity). It is, however, difficult to derive clear point-predictions from them in dynamic games.

An alternative class are outcome-based models in which the usual utility function is extended for elements that take the monetary payoff of other players into account. The most influential are Fehr and Schmidt (1999) and Bolton and Ockenfels (2000). Since both would yield very similar predictions, we focus only on one of the two in the following, namely the Fehr-Schmidt model. It basically assumes that people care about inequity, but to a different extent. In the two-player case the utility function of the Fehr-Schmidt model is given by

$$U_i(x) = x_i - \alpha_i \max\{x_j - x_i, 0\} - \gamma_i \max\{x_i - x_j, 0\} \quad (10)$$

with $i \in \{1, 2\}, i \neq j$, where $x = (x_1, x_2)$ denotes the vector of monetary payoffs, and we assume $\gamma_i \leq \alpha_i, 0 \leq \gamma_i \leq 1$. In this utility function the first term after x_i measures the utility loss that stems from inequity to i 's disadvantage and the last term measures the loss from advantages inequity.

For the sake of succinctness let us – following Fehr *et al.* (2001) – simply assume that there is a fraction $q \in [0, 1]$ of fair people in the population that exhibit $\alpha_i \geq \gamma_i > 0.5$, i. e. they have a willingness to pay in order to achieve equality. The rest of the population, i. e. $1 - q$ people, is purely selfish with $\alpha_i = \gamma_i = 0$. Consider a selfish principal that deliberates whether a wage level above the benchmark equilibrium would induce an agent to raise the chosen effort level. Assume for the moment that $p = 0$, i. e. there is no monitoring (that is what Fehr *et al.* (2001) call a trust contract). A fair agent who accepted a generous trust contract will choose an effort level that equalizes her own monetary payoff with the monetary payoff of the principal. Thus,

$$\Pi = \beta B(e) - w(e) - \mu M(p) = w(e) - \zeta C(e) = U \quad (11)$$

By using the implicit function theorem, one obtains

$$\frac{de}{dw} = \frac{2}{\beta B'(e) + \zeta C'(e)} \quad (12)$$

Since $de/dw > 0$ for all meaningful parametrizations, for a fair agent, e always increases with w . The important question, however, is whether the marginal effect on a principal's profit is greater than one. With a fraction q of fair agents, an increase of w by one unit increases average effort by

$$q \frac{de}{dw} = \frac{2q}{\beta B'(e) + \zeta C'(e)} \quad (13)$$

and the principal's profit by

$$\beta q \frac{de}{dw} = \frac{2\beta q}{\beta B'(e) + \zeta C'(e)} \quad (14)$$

If $\beta q de/dw = 2\beta q / (\beta B'(e) + \zeta C'(e)) \geq 1$, a higher wage level pays off even for a completely selfish principal. For an inequity averse principal a similar reasoning applies. It is, however, a bit more complicated, because inequity averse principals have to take the marginal effect of reciprocal behavior by the agent into account when deciding on w in order to be able to equate the two utility levels.

The arguments become a little less straightforward when incentive contracts that incorporate a positive monitoring probability are under investigation. Selfish principals would offer the optimal contract $\{p^*, w^*, e^*\}$ if all agents were selfish. If however some agents are fair, they run the risk that those agents might not accept their offers. Again, higher wage levels might, therefore, pay off for selfish principals according to similar conditions as for contracts without monitoring. In case of monitoring, principals can, however, also equate their payoff with the payoff of agents by setting the monitoring probability above equilibrium levels (where "equilibrium" refers to the standard solution with selfish principals and agents) and thereby creating fair contracts. If this would be perceived as fair by agents is, however, a question that cannot be answered in the Fehr-Schmidt framework. Depending on the parameters and especially on q , fairness equilibria with above-equilibrium monitoring may be Pareto-dominated by other fair contracts with equilibrium monitoring levels and above-equilibrium wages.

How will a fair principal decide? If the marginal effect of a wage increase on her profit is smaller than one (i. e. higher wages do not pay off in monetary terms), then she will choose the equilibrium monitoring probability to induce the equilibrium effort choice by the agent,

but a higher wage that will equally divide the surplus among the two. If the marginal effect is greater than one, the same intuition as for selfish principals applies.

Summing up the discussion on reciprocity and fairness, it is important to note that fairness motives are able to shift wages, efforts (and possibly monitoring levels) upwards in comparison to the standard solution, depending on the parameters of the game, the underlying functions and the fraction of fair people in the population. A short discussion, whether an upward shift is actually possible for our parameter choices is relegated to Section 4.

What can we say about the complementarity or substitutability of monitoring and pay in the presence of fair players? Since both standard equilibria and fairness equilibria (also without any monitoring) are possible, the answer is somehow unsatisfactory from a theoretical viewpoint, but reassuring for our claim that the question is ultimately empirical: Monitoring and pay might be both substitutes and complements, depending again upon parameters, specific functions and the fraction of fair people in the population as well as the conditions derived for the standard solution.

3 Existing empirical and experimental evidence

3.1 Empirical evidence

As already mentioned, there are rather few empirical studies investigating the effect of incentives and monitoring on performance. This is mainly due to some major econometric challenges. As Athey and Stern (1998) show, approaches that have been most commonly used in the literature can yield misleading results when one allows for complementarities between choice variables as well as unobserved factors that affect marginal costs and benefits of each individual choice. Although these issues can be dealt with theoretically, the requirements for field data are rather high.

One major challenge, e. g., is how to measure monitoring intensity. A common approach is to use the ratio of non-production to production employees (see, e. g., Gordon, 1990). However, this is only an approximation. Many of those included in the non-production category may have little or nothing to do with direct employee supervision. Other measures include self-reports and a measure of job autonomy that aggregates whether the employee has discretion over her work pace, whether she has to use a timekeeping system, and whether she has flexible working time.

Given the problems with field data, it is not very surprising that empirical results so far are rather ambiguous. Groshen and Krueger (1990) find evidence in favor of the traditional

efficiency wage model that predicts substitutability of pay and monitoring by looking at hospital employee data. The wages of staff nurses tend to fall with increasing supervision. Further, Rebitzer (1995) provides evidence that high levels of supervision are associated with lower wage levels by focusing on data from the petrochemical industry. Analyzing the data from a national survey, Kruse (1992) also finds a negative correlation between supervision and pay, and Arai (1994a,b) shows that higher wage premiums are associated with a larger fraction of autonomous jobs, where the level of autonomy is used as a proxy for monitoring intensity.

This does, however, not hold for the public sector, and in particular not for white-collar workers. Using survey data of employment conditions in the high-technology sector of a US state, Leonard (1987) shows that the traditional efficiency wage model is only weakly supported and that it fails to explain the high intra-industry dispersion. And even though Neal (1993) cannot provide direct evidence that wage premiums are not substitutes for monitoring activity as implied by traditional efficiency wage models, his empirical assessment does not support the derived hypothesis that inter-industry differences in monitoring contribute to inter-industry wage differentials. Finally, Gordon (1990) finds support for labor-discipline models, i. e. that pay and monitoring are complements.

The results of a recent survey by Minkler (2004) indicates that shirking in firms may not be as serious an issue as suggested by standard economic theory (see, e. g., Eaton and White, 1983). According to his results, moral and intrinsic motivation are very important determinants of workers' behavior. These forces are, however, more thoroughly investigated in experimental studies. Some of these studies are discussed in the following subsection.

3.2 Experimental evidence

Previous experiments on shirking and explicit incentives in a gift-exchange environment include Fehr *et al.* (1997), Fehr *et al.* (2001), Fehr and Gächter (2002), and based on more formal principal-agent models, Fehr *et al.* (1996), Keser and Willinger (2000), and Anderhub, Gächter and Königstein (2002).

The experimental designs of Fehr *et al.* (1996, 2001), and Fehr and Gächter (2002) share one peculiar and important design choice: They restrict the action space to variations of the fine in case of verified shirking instead of allowing for different degrees of monitoring. Legal constraints in real life may, however, lead to exactly the opposite. The fine may be fixed at some agreed upon level or might be implicitly given by the loss of rent when being dismissed, while the employer may have almost full decision autonomy regarding the intensity of supervision. In fact, penalty schemes are rare in practice on labor markets, with the exception of the penalty of getting fired.

The results of the above-mentioned experiments that incorporate an endogenous contract choice indicate that incentive contracts framed as bonus contracts are preferred over contracts framed as penalty contracts (for a discussion see also Luft, 1994). They also show that trust contracts, i. e. contracts without explicit incentives, are rarely chosen by the employer. Although in about one fourth of all incentives contracts the agent shirks, the studies provide mixed results concerning the average effort level under the different contract types. With an endogenous contract choice, effort is higher under incentive contracts, while effort is higher under trust contracts with exogenous contract choice. This observation is somewhat puzzling. If one assumes that the intentional choice of the contract design determines how the contract conditions are perceived - which is a logical implication of assuming that intentional kindness, i. e. high wages, leads to reciprocally kind reactions, i. e. a high effort choice - one will expect that choosing a trust contract is perceived as more kind than choosing an incentive contract. This, however, would imply higher effort levels under trust contracts with endogenous contract choice and, thus, contradicts the above observation. See also Bénabou and Tirole (2003) for a formal discussion on why explicit incentives may be counterproductive.

Keser and Willinger (2000) test a standard moral hazard model in the laboratory. In their experiment the agent can choose between two hidden actions that entail either low or high costs for the agent and on which the stochastic realization of either low or high gains for the principal depends. The principal offers an outcome-contingent contract, i. e. two wage levels, either of which is paid if low or high gains are realized. Note that this kind of contract closely resembles a bonus contract. Keser and Willinger observe that the offered contracts show the following features: The wage contingent on high gains is at least as high as the wage contingent on low gains, as it is required by theory. Agents do not have to risk a loss. And finally, the net profit of agents is not higher than the net profit of principals.

A further study on explicit incentives and contract design has been conducted by Anderhub *et al.* (2002). Principals offer a contract that consists of a wage, a profit share, and a non-binding desired effort level. The contracts are restricted to the space of linear contracts. However, incentive compatible contracts that induce efficiency are feasible and optimal. Agents have a piece-wise linear, convex effort-cost function that allows identifying a limited number of conditionally rational effort choices. The authors observe that principals offer incentive compatible return shares and ask for negative wages, i. e. entry fees. Agents often choose best reply efforts. According to Anderhub *et al.* deviations from the normative solutions can be explained by reciprocity. More generous contracts lead to a higher probability that agents act reciprocal, i. e. the deviations from conditionally rational effort choices are positively correlated with the surplus share. Agents reject unfair contracts and principals respond by offering contracts that

are fairer than predicted by the theoretical optimal contract design under the selfishness assumption. Consequently, earnings are less asymmetrically distributed than predicted. This indicates that vertical fairness concerns may influence contract design and incentives. However, participants have symmetric and complete surplus information, what may at least partly drive the results.

Let us complete the discussion of laboratory results with two papers investigating the effect of different productivities and effort costs. Hannan, Kagel and Moser (2002) extend the basic design of the gift exchange game (see Fehr *et al.*, 1993) by introducing low and high productivity firms. In their laboratory experiment workers do not provide more effort to lower productivity firms even though it is relatively more costly for these firms to offer higher wages. This is in contrast to the results of Gneezy (2004). Gneezy conducts a real-effort experiment in which the roles of employers and employees are assigned to MBA and undergraduate students, respectively. For solving mazes (the treatment conditions are two levels of difficulty: easy and hard) employees can be offered either 0, 5, or 10 dollars. Employers earn 1 dollar for each solved maze. In a third treatment using the easy mazes employers earn 3 dollars for each solved maze. Regardless of the treatment condition, each remuneration is chosen with equal probability by the employers, what may be due to the lack of experience with the task. There is a positive relation between wage and the number of solved mazes as well as between wage and the invested time. The employees' effort levels depend on the return level. They invest the same amount of time in the 1-dollar-per-maze treatment, but they reduce their effort in the 3-dollar-per-maze treatment. However, total earnings of the employers increase with the wage offer only in the 3-dollar-treatment. This supports Akerlof (1982)'s prediction that only when the return on employees' effort is sufficiently high, profits are increased by wages above the market-clearing level.

Finally, there is also a piece of evidence on monitoring and shirking from controlled field experiments. Nagin, Rebitzer, Sanders and Taylor (2002) investigate how employees of a call center company react to different exogenously given monitoring rates. They observe that a significant fraction of employees respond to a reduction in the perceived cost of opportunistic behavior by increased shirking. Contrariwise, employees with good outside options do not increase shirking by more than other employees when the rate of monitoring declines. Additionally, there also exists a significant number of employees who do not respond at all to variations in the monitoring intensity. This shows that the problem of shirking may indeed not be as serious as predicted by theory, but it is still prevalent and has to be taken care of by appropriate contract designs.

Table 1: Normative solution under the different treatment conditions

β	μ	p^*	w^*	e^*	$\Pi(p^*, w^*, e^*)$	$U(p^*, w^*, e^*)$	$\Pi^* + U^*$
12	30	0.41	10.90	2.53	6.42	6.88	13.30
12	20	0.57	13.97	3.79	8.73	6.58	15.31
16	30	0.63	24.46	5.99	16.60	9.79	26.39
16	20	0.88	31.74	8.99	22.06	4.79	26.85

4 Hypotheses and experimental design

For our experiment we use the following parametrization that fulfill the above conditions for pay and monitoring being complements in the theoretical solution. The benefit for the employer of an effort e is

$$\beta B(e) = \beta e^{2/3}, \quad (15)$$

the cost function for the employee of an effort e is given by

$$\zeta C(e) = \zeta e^{3/2} \quad \text{with } \zeta = 1, \quad (16)$$

and the cost of implementing a shirking detection probability p is

$$\mu M(p) = \mu p^2. \quad (17)$$

The minimal feasible effort level that an employee can exert if she is employed is $e = 0.1$. The minimal compensation an employer can offer is $w = 1$. The experimental treatment conditions are characterized by systematically varying the values of β and μ in a balanced 2×2 between-subjects design. More specifically, β will take either the value 12 or 16, and μ will take either the value 20 or 30. The motivation for the parameter variation is to picture different productivity and monitoring costs combinations that may be equivalent to different industry situations in the real-world. The normative solution of the shirking model assuming selfishness and using these parameters is presented in table 1.

Since the task is not easy, we cannot expect that participants will find the optimal contract immediately. Though, they should converge over time to one of the equilibria presented in Section 2. A money-maximizing employer should, however, always offer only contracts that are characterized by enforceable effort levels. Otherwise it is optimal for the employee to shirk.

Let us now briefly turn to the issue of fairness. The main task here is to check for our

parameter choices whether $\beta q^{de}/dw = 2\beta q/\beta B'(e)+\zeta C'(e) \geq 1$. If this is the case, even selfish principals will offer above-equilibrium wages to induce higher effort levels by agents. It can be shown³ that for our functional choices and β -values, $q \geq 0.4$ suffices to satisfy the condition for almost all effort levels.⁴ It is noteworthy that $q = 0.4$ is exactly the aggregated calibration result in Fehr and Schmidt (1999). We, therefore, expect to observe a positive effort-wage correlation in our data and non-minimal effort levels even without an investment in monitoring. Furthermore, if reciprocity is strong enough, i. e. the effort-wage slope steep enough, profits should be higher without monitoring than with monitoring for any given effort level.

However, as the employer might want to insure herself against exploitation (e. g., as a consequence of being let-down averse), he will monitor, but also offer a higher rent to induce non-shirking behavior of employees. This means that contracts will rather be characterized by lower than equilibrium monitoring and/or higher than equilibrium wages.

5 Experimental procedure

The experiment was conducted computerized using the software package zTree (Fischbacher, 1999) at the University of Innsbruck in January 2005. Participants were randomly recruited from the undergraduate population of the university. In total 124 students participated in the experiment. In any of the eight sessions that constitute one independent observation each either 18 or due to some no-show-ups 14 students participated. There were, however, always two employees more on the labor market than employers.

After arriving, the students were seated at screened computer terminals devided by blinds. The instructions were distributed and read aloud by one experimenter. They were framed in terms of a labor market in order to make the experiment less abstract and easier to understand for the participants. A specimen of the experimental instructions can be found in the Appendix. Then, participants were asked to answer a short control questionnaire to assess whether they understood the experimental rules and, in particular, the payoff determination. All participants had to pass this test. For being successful, they earned 3 €. Since it was possible to incur losses, the participants' attention was explicitly invited to this point. Participants had to agree to cover any losses immediately after the experiment.⁵ Before the first period, subjects were

³ Results are available from the authors upon request.

⁴ Notice that the marginal costs of effort provision $C'(e)$ is not constant in our framework, which leads to different results for different effort levels. With $q = 0.4$ the condition, however, holds for any $e \geq 1$.

⁵ No student rejected to participate due to this requirement. In case of losses – that in a few cases were substantial – they were indeed willing to pay their debts immediately.

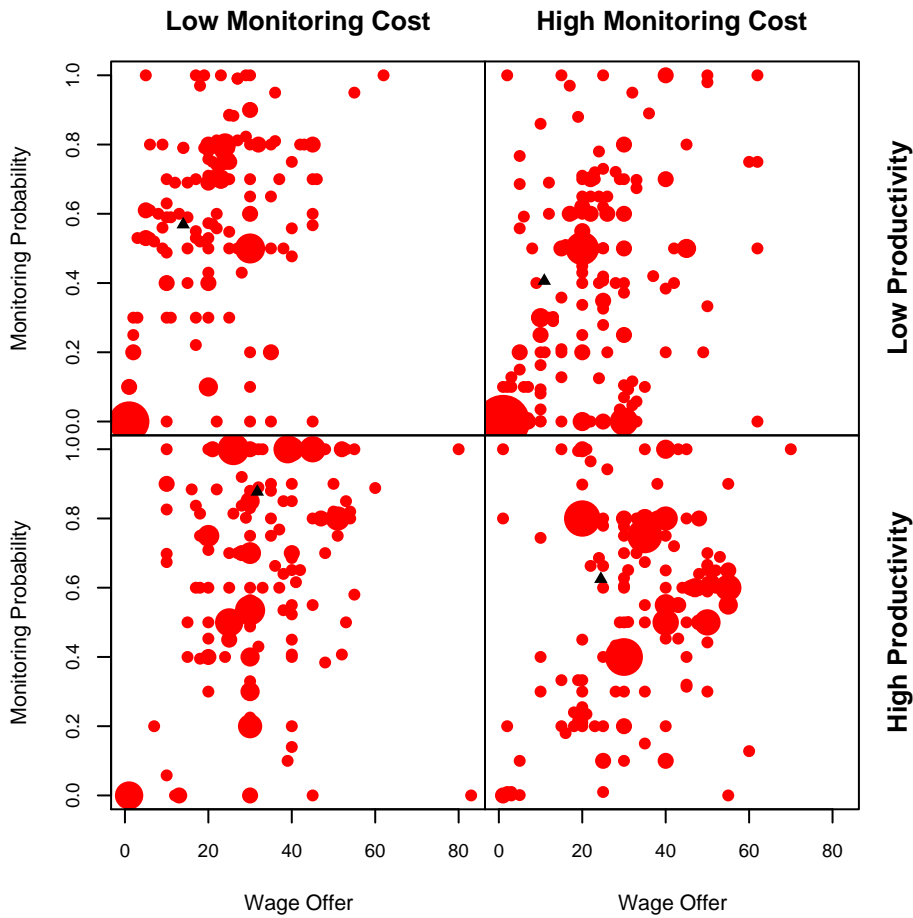
assigned to their roles. They kept their roles throughout the whole experiment. Yet, subjects were completely anonymous and not identifiable.

The sequence of actions within any period was: First, employers simultaneously decided about the offered contract, i. e. the level of monitoring p , the wage level \bar{w} to offer, and the desired effort level \tilde{e} . To facilitate the task, we, of course, restricted the parameter domain of variables for the participants. In particular we used $p \in [0, 1]$, $\bar{w} \in [1, 83]$, and e as well as $\tilde{e} \in [0.1, 12]$. After the contract offer phase all offers were made public to the workers in a random order. One after another they could, then, choose to accept any still available contract or to reject all standing offers. As soon as one worker accepted a contract offer, the offer was not available for any other worker. This posted offer market stage ended either after all contracts were accepted or all unemployed workers had rejected the still available contracts. Subsequently, all employed workers chose their effort level by moving a slider on a continuous scale. Finally, uncertainty about monitoring was resolved, and payoffs were determined according to the actions of the participants and the chance move.

In order to facilitate calculations the computer program offered the participants the following help: Employers were provided with a calculator for the monitoring cost, the (desired) effort costs, and gross profits in case of both compliance and shirking. Thus, subjects could design their contracts by comparing different contractual conditions and choose the values that suited them best. By providing the gross profits in case of compliance and shirking, we made sure that subjects were fully aware of possible gains and losses. Employees were provided with a calculator for their effort costs as well as their own and their employer's earnings, given the accepted contract and the self-defined effort level. Employees could, therefore, compare the results of exerting various effort levels and choose the preferred. We believe that supplying our subjects with these pieces of information was important to obtain more experienced choices and a faster convergence. The alternative of playing more rounds would have been less desirable in our view.

After 15 periods the experiment ended, and participants were paid in private. A typical session took less than 90 minutes in total.

Since there are substantial differences in earning opportunities measured in experimental points between the four treatments both at and off equilibrium, we used different exchange rates. In the treatment with low productivity and high monitoring cost (henceforth, LPHM) experimental points were converted at an exchange rate of 7 points per euro. In the treatment with low productivity and low monitoring cost (LPLM) the conversion rate was 8 points per euro; and in both high productivity treatments (i. e. HPLM and HPHM) the exchange rate was set to 13 points per euro. Consequently, the sum of earnings in the equilibrium was approxi-



Note: The size of the bubbles is proportional to the number of observations represented. The black triangle denotes the optimal contract according to the normative solution under the selfishness assumption.

Figure 1: Distribution of contracts in the monitoring probability and wage offer space

mately the same in all treatments. Average earnings in the experiment were € 8.64 (standard deviation: 10.57).

6 Experimental results

Let us first discuss contract offers. Figure 1 shows the distribution of contract offers in monitoring probability and wage space. The black triangles denote the optimal contract according to the normative solution under the selfishness assumption. It is immediately clear from the figure that there is considerable variance in the data. However, in the low monitoring cost treatments (LPLM, HPLM) contracts are rather clustered around the optimal contract. In the high monitoring cost treatments (LPHM, HPHM) offered wages seem to be too high. It is further

Table 2: Frequency of accepted contracts without monitoring

Productivity	Low Monitoring Cost		High Monitoring Cost	
	Low	High	Low	High
Monitoring >0	159	192	152	200
Probability 0	21	16	51	4

Table 3: SUR estimation of wage offer and monitoring probability

Wage Equation					
Coefficient		Estimate	Std.Error	t-value	p-value
Intercept		21.028	1.031	20.390	<0.001
High Productivity		9.787	1.405	6.964	<0.001
High Monitoring Cost		-1.509	1.405	-1.074	0.283
High Productivity : Monitoring Cost		4.942	1.949	2.535	0.011
Multiple R-Squared		0.172			
Adjusted R-Squared		0.169			
Monitoring Probability Equation					
Coefficient		Estimate	Std.Error	t-value	p-value
Intercept		0.550	0.022	25.258	<0.001
High Productivity		0.079	0.030	2.673	0.008
High Monitoring Cost		-0.209	0.030	-7.033	<0.001
High Productivity : Monitoring Cost		0.138	0.041	3.363	0.001
Multiple R-Squared		0.124			
Adjusted R-Squared		0.120			
Correlation of Residuals		Wage	Monitoring		
Wage		1.000	0.406		
Monitoring		0.406	1.000		

striking that in the low productivity conditions (LPHM, LPLM), we find a substantial number of contracts that include no monitoring (see also table 2). Yet, the clear majority of the contracts are incentive contracts.

As wage and monitoring probability are chosen simultaneously, we need to account for this when we want to assess whether participants react to the treatment conditions and whether this reaction is in the direction predicted by the model. We, therefore, estimate an equation system on wage and monitoring probability by applying the seemingly unrelated regression (SUR) approach. The results are presented in table 3.

Table 4: Random coefficient estimation on wage offers

Coefficient	Estimate	Std.Error	t-value	p-value
Intercept	13.255	3.353	3.954	<0.001
Monitoring Probability	14.258	3.305	4.314	<0.001
High Productivity	9.945	3.431	2.899	0.044
High Monitoring Cost	2.223	3.524	0.631	0.562
High Productivity : Monitoring Cost	0.488	4.847	0.101	0.925
Random Coefficients	On Level	Std.Deviation	Correlation	
Intercept	Matching Group	0.023		
Intercept	Subjects	16.674	Intercept	
Monitoring Probability	Subjects	20.399	-0.872	
Residual		6.664		

The regression results provide clear evidence that, on average, participants reacted to the treatment variations. Note again that this is a between-subject comparison. Further, contract choices are almost always adjusted in the right direction, i. e. high productivity leads to higher wages and a higher monitoring probability, and high monitoring cost leads to less intense monitoring and lower wages. However, the adjustment of wages to increasing monitoring costs is not significant in the low productivity treatments. It is nevertheless noteworthy that we obtain significant results for all other treatment dummies despite the large variance in the data that is apparent from figure 1. Additionally, wages are too high compared to the normative solution under the assumption of selfishness in all treatments with the exception of HPLM, in which wages are just at the right level. This is also the only treatment in which average monitoring probabilities are too low, otherwise they are rather close to the equilibrium (see also figure 6). The observed distribution of wages and monitoring probabilities is, thus, a first indication that employers may rely on or want to induce reciprocal behavior on the side of the employees.

Having established that the between treatment variation in contracts is qualitatively as predicted by the model, we need to verify whether the within treatment variation is also in line with the normative prediction. To this end, we run a random coefficients regression on wage offers in which we include the monitoring probability and treatment dummies. A likelihood ratio test indicates that we need to include the monitoring probability as a random coefficient on the subject level, i. e. there is considerable heterogeneity in the correlation between monitoring probability and wage offer between subjects. We further include a random intercept on the matching group and subjects levels. This procedure allows us to control for

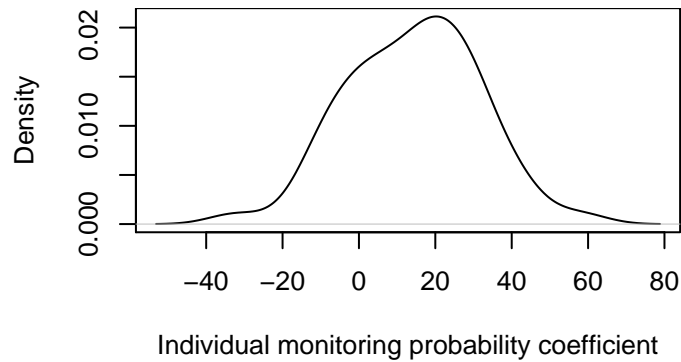


Figure 2: Distribution of individual monitoring probability coefficients in the regression on wage offers

the repeated measurement and the within matching group correlation. We further control for treatment-specific heteroskedasticity. The regression results show that there is, on average, a highly significant positive correlation between wages and monitoring probability (see table 4). Figure 2 where we show the distribution of individual coefficients for the monitoring probability reveals that due to the considerable heterogeneity a few subjects (less than one quarter) actually implemented a series of contracts trying to use monitoring and pay as substitutes. We see, however, that most participants indeed used pay and monitoring as complements. By applying further likelihood ratio tests, we checked whether the correlation is different in the various treatments. There is, however, no such evidence in the data.

Wage offer and monitoring probability jointly determine the enforceable effort level. If the desired effort level is equal to or lower than the enforceable effort level, the respective employee should exert the desired effort level. Otherwise, full shirking maximizes her payoff. Figure 3 presents a scatter-plot of desired, enforceable and actual effort levels. The following observations emerge from the plot: First, most desired effort levels are higher than the corresponding enforceable effort levels. In fact, in only 199 out of 795 accepted contracts the desired effort level is enforceable. Second, given the contracts, employees either fully shirk, partly shirk, exert just the desired effort, or exert even a higher effort level. While fully shirking and exerting just the desired effort level can be easily explained, the other two options question either the conditional rationality of employees or might be the result of reciprocity considerations. Yet, despite of the heterogeneity, it is easy to discern from figure 3 that a majority of the decisions is clustered either around the predicted 45 degree line or at the zero-effort level.

According to a logistic regression on shirking frequencies (see table 5) where we control for repeated measurement, we observe that there is less shirking in the high productivity treat-

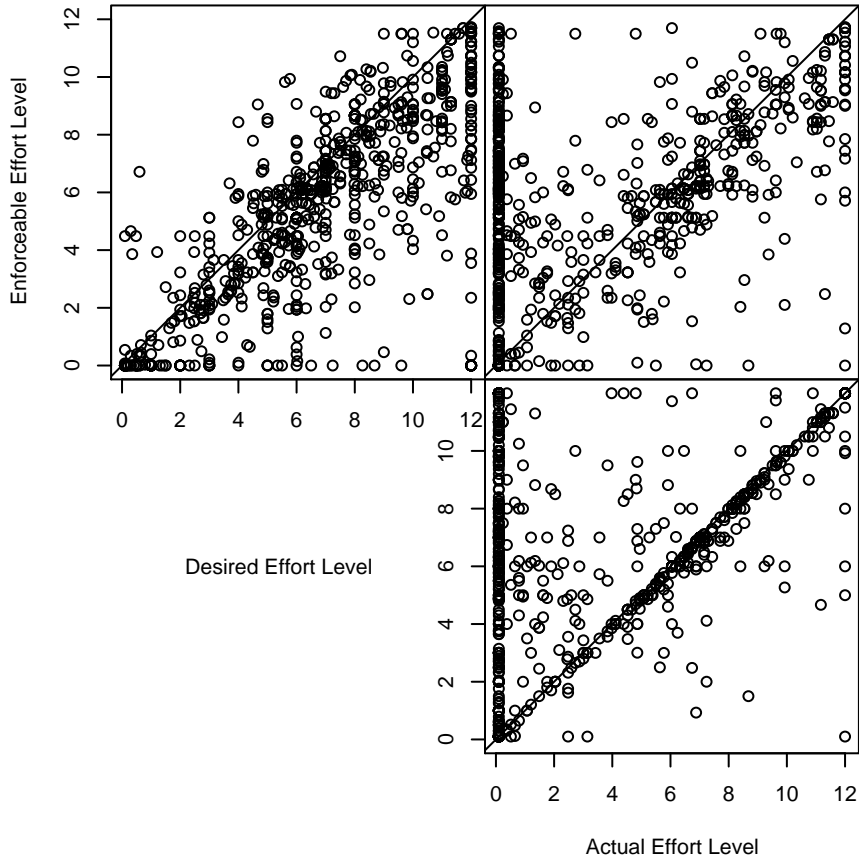
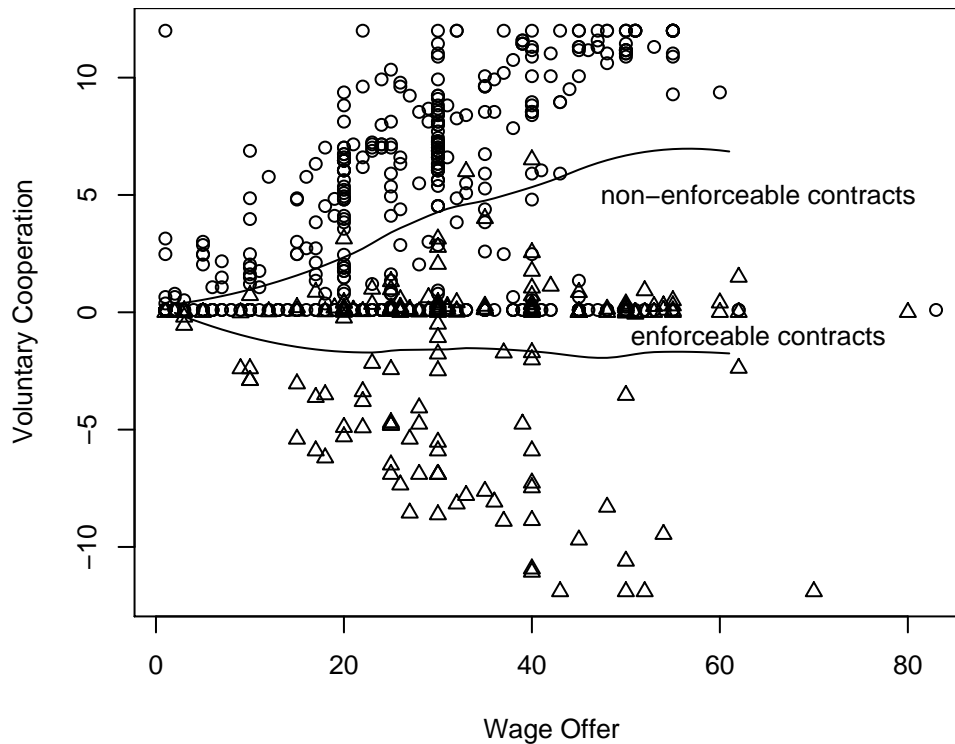


Figure 3: Desired, enforceable and actual effort levels

Table 5: Frequency of shirking decisions

Desired Effort Level	Low Productivity		High Productivity	
	Shirking	No Shirking	Shirking	No Shirking
Enforceable	40	36	32	91
Not Enforceable	206	101	153	136

ments (likelihood ratio test; $p = 0.014$) and – as one should expect – when the desired effort level is enforceable ($p < 0.001$). Further, a logistic regression on shirking frequencies where we restrict the sample to include only the non-enforceable contracts reveals that the probability of shirking decreases with the monitoring probability ($p < 0.001$) and the offered rent ($p < 0.001$), i. e. wage minus cost of desired effort. The last observation is another clear evidence for the existence of reciprocity in our experiment.



Note: Circles denote the amount of voluntary cooperation under non-enforceable contracts while triangles denote the amount of voluntary cooperation under enforceable contracts. Under non-enforceable contracts the voluntary cooperation cannot be negative. The lines show local averages obtained by non-parametric local fitting.

Figure 4: Voluntary cooperation under enforceable and non-enforceable contracts

To further explore the impact of reciprocity considerations on effort choices we define the amount of voluntary cooperation of an employee as the difference between her actual effort choice and the conditional optimal effort choice given the accepted contract. In figure 4 we show all data points where we distinguish between enforceable and non-enforceable contracts. With non-enforceable contracts the amount of voluntary cooperation cannot be negative. Although the data reveals a huge heterogeneity in decisions we observe a clear positive correlation of wages and voluntary cooperation under non-enforceable contracts. Under enforceable contracts there is no such correlation. We can, however, observe that agents shirk even under enforceable contracts (in 72 out of 199 cases).

A random coefficient regression on the amount of voluntary cooperation that includes random effects for the matching group and the individual subjects, treatment dummies, a dummy for zero monitoring, a dummy for enforceable contracts, the wage that had to be modelled as a random coefficient according to a likelihood ratio test, and the interaction between wage and zero monitoring and enforceable contracts, respectively – where we control for the repeated

Table 6: Random coefficient regression on the amount of voluntary cooperation

Coefficient	Estimate	Std.Error	t-value	p-value
Intercept	-0.079	0.396	-0.199	0.842
High Productivity	0.612	0.357	1.713	0.147
High Monitoring Cost	0.323	0.336	0.959	0.382
No Monitoring	0.254	0.506	0.502	0.616
Enforceable Contract	-1.598	0.647	-2.469	0.014
Wage	0.128	0.014	9.224	<0.001
Wage: No Monitoring	-0.115	0.025	-4.644	<0.001
Wage: Enforceable Contract	-0.146	0.020	-7.323	<0.001

Random Coefficients	On Level	Std.Deviation	Correlation
Intercept	Matching Group	<0.001	
Intercept	Subjects	<0.001	Intercept
Wage	Subjects	0.051	-0.211
Residual		2.980	

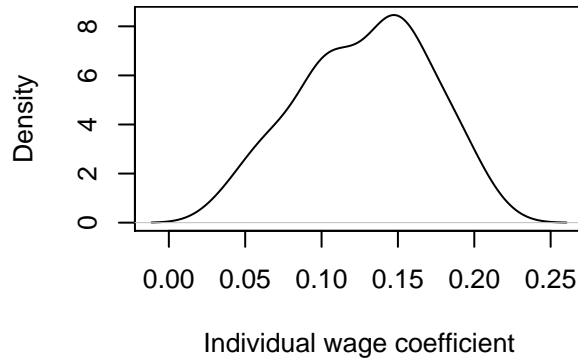


Figure 5: Distribution of individual wage coefficients in the regression on voluntary cooperation

measurement and heteroskedasticity – reveals the following (see table 6): First, we observe no treatment effects. We also tested for possible interaction effects of the treatments with the offered wage on the amount of voluntary cooperation. Yet, none of the likelihood ratio tests suggested a significant interaction effect. Second, there is a significant positive impact of wage on the amount of voluntary cooperation. A wage increase of 10 points leads to a voluntary increase in the effort level of about 1.3 units. The standard deviation of the random coefficient for the wage offer reveals that participants react rather heterogenously. Individual wage coefficients vary between 0.037 and 0.211 (cf. figure 5). Third, the positive effect of high wages

is offset when the contract was enforceable. Under this condition there is no wage–voluntary cooperation correlation. Due to some shirking we observe, however, some negative voluntary cooperation independent of the wage. The amount of voluntary cooperation cannot be negative if contracts are non-enforceable as any non-minimal effort is considered a voluntary cooperation. As a result, given the same wage actual effort choices are higher when contracts are enforceable. Finally, if there is no monitoring the positive effect of wages on voluntary cooperation are offset, too. Without monitoring we almost always observe minimal effort choices regardless of the offered wage. To induce any non-minimal effort choices principals have to set a positive monitoring probability.

As the marginal effect of a wage increase on the principal’s benefits is below one for all efforts below 1.12 and 2.69 for the low and high productivity treatment, respectively, the principals fare best by implementing contracts that adhere to the normative solution of the shirking model. Relying on reciprocity does not pay. Given the same wage, average per round earnings are 5.6 (non-enforceable desired effort, $p = 0.019$) to 6.0 points (enforceable desired effort, $p = 0.032$) higher when the employer decides to monitor. This, of course, explains why incentive contracts are predominant.

Let us conclude the presentation of the experimental results with an analysis of behavior over time. Figure 6 shows the mean offered rent, monitoring probability and effort for all 15 periods. The offered rents are too high during the first periods. However, except for the HPLM treatment, they sharply decline and stabilize after period four. While the mean offered rents are very close to the normative prediction afterwards in the high productivity treatments, they are too low at the end of the experiment in the low productivity treatments.

The corresponding mean monitoring probabilities start at about the right level, with the exception of the HPLM treatment where the monitoring intensity is too low. In this treatment the monitoring probability slightly increases over time, but stays below the normative prediction. In the LPHM treatment, the mean monitoring probability stays stable until period ten. Afterwards it decreases considerably below the normative prediction. In the two other treatments the monitoring probability stays rather stable over the whole experiment. Though, it is slightly below the normative prediction during the last few periods.

The number of the resulting enforceable desired effort levels in each period is given in table 7. Although the number seems to decline over time, the trend is not significant on conventional levels (Chi-squared test for trend in proportions, $p = 0.056$).

Finally, actual effort levels are close to the normative prediction during the first half of the experiment in the low productivity treatments, but below afterwards due to the decrease in offered rent and monitoring probability. Efforts in the HPHM treatment are rather volatile

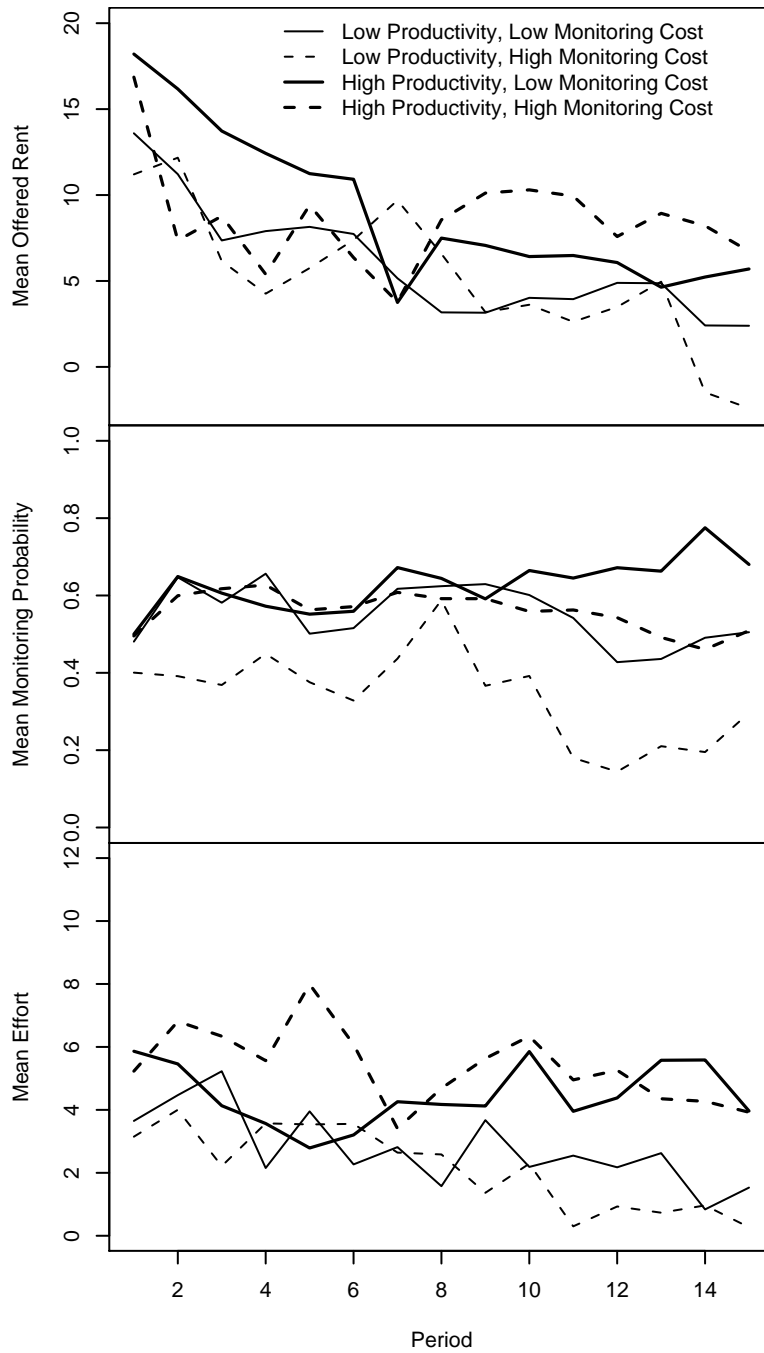


Figure 6: Mean offered rent, monitoring probability, and effort over time

Table 7: Number of (non-)enforceable desired effort levels over time

Effort	Period														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
non-enforceable	33	32	42	40	44	45	44	39	41	45	39	43	43	42	39
enforceable	21	22	12	14	10	9	10	15	13	9	15	11	11	12	15

but most times close to the normative prediction. There is a slight downward trend that leads to efforts being below the normative prediction at the end. In HPLM, the efforts are always considerably below the normative prediction, which is mainly due to the always too low monitoring probability and the too low offered rents in later periods.

The described evolution of contracts and effort choices leads to no significant trend for employers' earnings ($p = 0.245$; Wald test, using a random effects regression on earnings in euro), but to a slight decrease in employees' earnings over time (about 3 euro-cents per period; $p < 0.001$).

Although we again observe some variance, we can conclude that the overall adjustment processes go in the right direction with only a few exceptions. Especially the predicted treatment differences become obvious over the course of the experiment.

7 Discussion

We set out on testing a shirking model in which the monitoring intensity arises endogenously. Wage level, monitoring intensity and consequently the desired enforceable effort level are jointly determined by the maximization problem of the firm. As a result, monitoring and pay should be complements.

The question was, whether behavioral regularities, i. e. reciprocal behavior, might be able to change the nature of these strategic complementarities such that they are perceived and used as substitutes, as suggested by fairness models.

To this end, we designed an experiment that has two distinctive features. First, in contrast to earlier experiments that investigate shirking models the monitoring intensity in our experiment is endogenous. Instead of fixing a fine – that should be maximal (what is actually also observed in those earlier experiments), given an exogenously fixed shirking detection probability – participants have to choose a monitoring probability together with a wage level and a desired effort. Second, feasible efforts are not restricted to a small set of discrete levels, but are chosen on a continuous scale. Using a continuous effort is crucial for the prediction of the shirking model. If effort would be a discrete variable, say taking on only the values low or high, monitoring and pay would become substitutes (see Allgulin and Ellingsen, 2002). Such an oversimplification would, thus, completely alter the predictions of the model and lead to opposite guidelines for contract design.

Though contracts are not as predicted by the normative solution of the shirking model and do not converge perfectly to the prediction over the course of the experiment, the between treatment variation is qualitatively as predicted: High profitability leads to higher wages and

higher monitoring intensities, and high monitoring costs lead to a reduction in both. However, wages are generally too high and the monitoring probability is too low, indicating that employers pay higher rents than implied by the model. Analyzing the within treatment variation, we find a clear and highly significant positive correlation between monitoring and pay. Thus, qualitatively the predictions of the shirking model are confirmed: Monitoring and pay are complements. Given the relatively high variance in our data, obtaining statistically significant treatment differences can be viewed as a strong indication that the theoretical predictions are corroborated experimentally. One intuitive explanation of why subjects in the role of the employers might view monitoring and pay as complements is that they may have the feeling that if they pay well, they have the moral right to check the quality of the work. Such a reasoning – which is against the conventional wisdom – would immediately imply that supervision and wages are complements.

Nevertheless, we also observe clear evidence for the existence of reciprocity. In a labor market context, reciprocity is usually characterized by a positive wage- or rent-effort correlation, which is the result of mutual gift-exchange (Akerlof, 1982). As the reciprocity idea implies that intentions matter, we should observe less shirking with higher rents and a higher rent-effort slope when contracts comprise zero monitoring. While the first effect is clearly reflected in our data, we find no support for the second implication of reciprocity. On the one hand, shirking rates decrease with higher rents, and the amount of voluntary cooperation significantly increases with a higher rent. On the other hand, the latter is not true when there is zero monitoring. Although this last observation is somehow disturbing and difficult to explain in the framework of common fairness models, it was also observed by Fehr *et al.* (1996, 2001) and Fehr and Gächter (2002). Consequently, one result of the shirking model, namely that there must be a positive level of monitoring in order to induce any effort, certainly prevails in an experimental test.

Accordingly, at least for the parameters and functions chosen in our experiment relying on reciprocity does not pay off. The employer's earnings are at least as high when she designed a contract with enforceable effort as when she invested the same amount of money in the labor relationship but offered a higher rent and monitors less.

Finally, as already Kirchler, Fehr and Evans (1996) observed, there is considerable heterogeneity among participants. The rent-effort relation is by no means the same for all participants. Even though the correlation is on average significantly positive, there is a substantial number of participants who do not or very little react to the rent, while there are also participants who react very strongly to changes in the offered rent.

In summary, the answer to our question whether the social norm of reciprocity is strong

enough to change the complementary character of monitoring and pay as instruments of contract design is negative. It can, of course, only be supported for the analyzed parametrization in our experiment. We believe, however, that the four treatment conditions that we put to a test span a relatively large space of feasible parameter combinations. Note further that other experiments have shown that the positive effects of reciprocity are strongest only for specific environments in which, for instance, extremely high productivities and complete certainty about payoff consequences of one's actions prevail. As these conditions are not (always) met in the real world, it seems important to use different set-ups to be able to learn more about such important behavioral regularities like reciprocity on labor markets.

This brings us to the important question of where we believe to find future paths of research in that area. Our experiment demonstrates that, while the existence of reciprocity is an established fact in economics, we still require more evidence on its important interactions with incentives. When some experiments show that specific incentives may crowd out reciprocal motives and others like ours that for many people a low-intense incentive is necessary to evoke reciprocity, it becomes an obvious concern to be able to distinguish environments with regard to this issue. Note that experiments with an endogenous choice between trust contracts and incentives contracts, in which subjects have a preference for incentive contracts, confirm our results. Even if economics is still quite agnostic with regard to the specific interaction effects between incentives and reciprocity, our experiments have added another important piece of evidence required to understand the effects of incentives: In labor relations between an employer and an employee, a low-intense incentive goes a long way. Even if it would be optimal to establish strong monitoring mechanisms (e. g., when they are almost costless), it may be a good idea to cut down monitoring to a lower intensity and provide some contractual freedom. However, full trust and zero monitoring does also not work, on average.

8 Appendix

This is the translation of the instructions for the HPLM treatment. Instructions for the other treatments are identical except for the parameter values and the exchange rate.

8.1 Instructions

This experiment analyses economic behavior. During the experiment you and the other participants will make decisions, and you will earn money. The amount of money you earn depends on your own decisions as well as on the decisions of the other participants and is determined

by the rules of the game that will be explained in the following in detail. At the end of the experiment your total profit will be privately paid to you in cash. If you have any questions after reading the instructions please raise your hand. One of the experimenters will come to you and answer your questions privately. All participants receive identical instructions.

8.2 Types of participants

In the experiment, there are two types of participants: participants A and participants B. It is most convenient for you if you view participants of type A as employers and participants of type B as employees or workers. The assignment to the two roles is completely random. You will learn your role on your screen at the beginning of the experiment. You will remain in your role throughout the entire experiment.

8.3 Earnings

At the beginning of the experiment you receive 3 euro, contingent on answering a questionnaire about the rules of the experiment correctly. During the experiment you will earn money by accumulating points. The accumulated points will be converted to euros at the end of the experiment. The conversion rate is: 1 euro = 13 points. At the end of the experiment all period profits will be added up and paid to you privately and in cash. In case you make losses in the experiment – which is unlikely but possible – you have to pay your losses to the experimenters. If you prefer not to participate in the experiment under these rules, we would ask you to tell us now and to leave the room.

8.4 Duration

The whole experiment will last for about 90 minutes. It is divided into 15 periods. In each period you will have to make decisions on the computer.

8.5 Anonymity

You will not learn the identity of the participants you are going to interact with, neither during nor after the experiment. Other participants will not learn about your role, your decisions and how much you earned. It is not allowed to talk during the experiment. You are not allowed to use other functions of the computer than the experimental program. Communication with others than the experimenters or manipulations on the computer will lead to you being expelled from the experiment.

8.6 Overview of course of action during the experiment

In each period of the experiment, an employer and an employee can conclude a trade. There are 8 employers and 10 employees in the market. Your role remains unchanged throughout the whole experiment. In each period, the course of action follows the same procedure: Each of the 15 periods starts with an offer phase. During this phase the employers have to submit offers that can, then, be accepted by employees. An offer comprises the following three items whose consequences will be described in greater detail later: A monitoring probability in the interval $[0, 1]$ (with a maximum of three digits after the comma); a wage in the interval $[1, 83]$ (only integer numbers); and a desired effort (=performance) in the interval $[0.1, 12]$ (with a maximum of three digits after the comma). For inserting this information the following screen will appear.

The screenshot shows a web-based interface for an experiment. At the top, it indicates the current period is '1 von 15' and the remaining time is '86' seconds. The main area is divided into sections. The first section states 'Sie haben die Rolle des Arbeitgebers. Wählen Sie bitte die Bedingungen Ihres Vertrages.' Below this, there is a section for calculating the costs of the monitoring probability, with a slider set to 0.0 and a 'Berechnen' button. The second section is for calculating the benefit of a desired effort, with a slider set to 0.1 and another 'Berechnen' button. At the bottom, there are three input fields for 'Kontrollwahrscheinlichkeit', 'Lohn', and 'Gewünschter Effort', each with a blue box. An 'OK' button is located at the bottom right.

The upper part of the screen will allow you to calculate two important values: First, you can calculate the costs of the monitoring probability contingent on your chosen values (a higher monitoring probability causes higher costs for the employer). To conduct the calculation you just have to pull the slider to the desired position and press the "Calculate"-button. Then, the costs in points will be shown.

Below, you can calculate how many points a desired effort would be worth for you (a higher effort increases the profit of the employer) and which costs a desired effort would cause for the employee who accepts your offer (a higher effort increases the costs for the employee). You can use the slider and the "Calculate"-button as often as you wish. When you finally inserted the values in the three fields and pressed the OK-button, your contract offer is valid.

Offers of employers are public and can be seen by all employees. Employers cannot see the offers of other employers, however. All employers have to submit an offer. Employers can only submit one offer each period. Each participant can in every period conclude a maximum of one trade. Since there are 8 employers and 10 employees, at least 2 employees remain without a trade in any period.

After all employers will have submitted their offers, the acceptance phase will start. In this phase, employees will be able to accept offers submitted by employers. During this phase, employees see the screen with the contract offer in consecutive but randomly determined order. In each period this random order will be newly determined. Each employee can, then, in the course of 20 seconds decide which contract offer to accept, or to decline all standing offers. Accepted offers will be deleted from the screen for the subsequent employees. During this phase, employees will see the following screen:

Periode
1 von 3
Verbleibende Zeit [sec]: 3

Entscheiden Sie sich bitte für eines der Angebote. Sie können auch alle Vertragsangebote ablehnen.

Gewünschter Effort:	Lohn:	Kontrollwahrscheinlichkeit:
2.000	7.0	0.500
12.000	40.0	1.000

You can activate single contract offers by clicking on them (they will, then, be shaded in blue) and accept them by clicking the OK-button. If you do not want to accept any of the offers, please click on the "No contract"-button. After 20 seconds the possibility to accept offers ceases for the employee. If you have not accepted an offer by then, you will remain without a trade in this period.

It is straightforward to view the offer and acceptance phases as a labor market, on which employers make offers on work contracts (consisting of the monitoring probability, a wage and a desired effort). Employees can either accept one of these offers or decline all of them.

After the acceptance phase all employees who have concluded a trade have to submit an actual effort level to the employer they have contracted with. For the employees the desired effort level by the employers are NOT binding. You will see the following screen:

The screenshot shows a software interface with a yellow border. At the top left, it says "Periode" and "1 von 15". At the top right, it says "Verbleibende Zeit [sec]: 26". The main area contains the following text and elements:

Sie haben sich für folgenden Vertrag entschieden:

Lohn:	1
Gewünschter Effort:	0.100
Kontrollwahrscheinlichkeit:	0.000

Wählen Sie nun den Effort, den Sie tatsächlich leisten möchten.

Effort: 0.1 12

Bevor Sie sich endgültig entscheiden, können Sie die Effortkosten und Ihre Periodenauszahlung bei dem gewählten Effort berechnen lassen.

At the bottom right of the main area is a red button labeled "Berechnen". At the bottom right of the entire interface is a red button labeled "OK".

Employees can also use a slider and a "Calculate"-button on their screen to calculate the costs of effort, their period profit and the period profit of the employer for the chosen effort level. Again, the "Calculate"-button can be used as often as wished. After the OK-button will finally have been clicked, the chosen effort will, then, be valid. After all employees have chosen their efforts, period profits are determined and you will learn all important values as well as your period profit on a separate screen. Then, a new period starts.

8.7 How profits are calculated

8.7.1 Profits of employers

If an employer has not concluded a trade in a period (because nobody accepted it), she will earn a profit of 0 points in that period. If an offer has been accepted, the profit depends on the wage, the monitoring probability and the actual effort determined by the employee. The profit will be determined according to the following formulae:

1st possibility: The random mechanism has determined that there is no monitoring, or the actual effort was at least as high as the desired effort: Profit employer = $16 \times (\text{actual effort})^{(2/3)} - 20 \times (\text{monitoring probability})^2 - \text{wage}$

2nd possibility: The random mechanism has determined that monitoring takes place and the actual effort was lower than the desired effort: Profit employer = $16 \times (\text{actual effort})^{(2/3)} - 20 \times (\text{monitoring probability})^2 - 1$

As it is apparent from the formulae, the profit of the employer is the higher, the higher the actual effort determined by the employee, the lower the chosen wage and the lower the chosen monitoring probability.

Note that you do not have to use the formula to calculate profits! The slider and the "Calculate"-button helps you to decide without having to do the calculations for yourself.

8.7.2 Profits of employees

If an employee has not concluded a trade in a period, she will earn a profit of 0 points in that period.

If the actually chosen effort is at least as high as the desired effort submitted by the employer, or if the random mechanism decided not to monitor, then: Profit employee = $\text{wage} - (\text{actual effort})^{(3/2)}$

If the actually chosen effort is lower than the desired effort submitted by the employer and the random mechanism decided to monitor, then: Profit employee = $1 - (\text{actual effort})^{(3/2)}$

Hence, the computer decides randomly according to the monitoring probability that is chosen by the employer whether to monitor or not. If the actual effort is at least as high as the desired effort, monitoring has, of course, no consequences. In case the actual effort is smaller than the desired effort, the profit depends on whether the random mechanism decides to monitor or not.

Again, note that you do not have to use the formula to calculate profits! The slider and the "Calculate"-button helps you to decide without having to do the calculations for yourself.

Profits of all employers and employees are determined in the same way. Each employer can, therefore, calculate the profit of the employee with whom she concluded a trade, and each employee can calculate the profit of the employer with whom she concluded a trade.

After each period you learn as employer: the actual effort chosen by the employee (in case you concluded a trade) and your period profit. After each period you learn as employee: whether you have been monitored or not (in case you concluded a trade) and your period profit.

Note finally that both employers and employees can also make losses in each of the 15 periods. You have to cover these losses out of the 3 euros you will earn at the beginning of the experiment or out of profits in other periods. Your round profits will be added up over the 15 rounds and paid to you at the end of the experiment privately and in cash.

8.8 Questionnaire

Question 1: The offer of an employer was not accepted by any of the employees. What is her profit in this round?

Question 2: An employee has not accepted any offer. What is her profit in this round?

Question 3: An employer chose a monitoring probability of 0 and a wage of 30. The (actual effort)^(3/2) chosen by the employee was 2. What is the profit of the employee in this period?

Question 4: An employer chose a monitoring probability of 0.5, a wage of 20 and a desired effort of 5. The actual effort chosen by the employee was 4. The random mechanism chose to monitor the employee. What is the profit of the employee in this period? (note: $4^{(3/2)} = 8$)

References

- AKERLOF, G. A. (1982). Labor contracts as partial gift exchange. *Quarterly Journal of Economics*, **97**, 543–569.
- ALLGULIN, M. and ELLINGSEN, T. (2002). Monitoring and Pay. *Journal of Labor Economics*, **20(2)**, 201–216.
- ANDERHUB, V., GÄCHTER, S. and KÖNIGSTEIN, M. (2002). Efficient contracting and fair play in a simple principal-agent experiment. *Experimental Economics*, **5**, 5–27.
- ARAI, M. (1994a). Compensating Wage Differentials versus Efficiency Wages: An Empirical Study of Job Autonomy and Wages. *Industrial Relations*, **33**, 249–261.

- (1994b). An Empirical Analysis of Wage Dispersion and Efficiency Wages. *Scandinavian Journal of Economics*, **96**, 31–50.
- ATHEY, S. and STERN, S. (1998). An Empirical Framework for Testing Theories about Complementarity in Organizational Design. *Working Paper 6600*, National Bureau of Economic Research.
- BÉNABOU, R. and TIROLE, J. (2003). Intrinsic and extrinsic motivation. *Review of Economic Studies*, **70**, 489–520.
- BOLTON, G. and OCKENFELS, A. (2000). A theory of equity, reciprocity and competition. *American Economic Review*, **100**, 166–193.
- CAHUC, P. and ZYLBERBERG, A. (2003). *Labor Economics*. Cambridge, Massachusetts: MIT Press.
- CHANG, J. and LAI, C. (1999). Carrots or sticks? A social custom viewpoint on worker effort. *European Journal of Political Economy*, **15**, 297–310.
- DEMOUGIN, D. and FLUET, C. (2001). Monitoring versus incentives. *European Economic Review*, **45**, 1741–1764.
- EATON, C. and WHITE, W. D. (1983). The Economy of High Wages: An Agency Problem. *Economica*, **50**, 175–181.
- FEHR, E. and GÄCHTER, S. (2002). Do incentive contracts crowd out voluntary cooperation. *Working Paper 34*, Institute for Empirical Research in Economics, University of Zurich.
- FEHR, E., GÄCHTER, S. and KIRCHSTEIGER, G. (1997). Reciprocity as a contract enforcement device: Experimental evidence. *Econometrica*, **65**(4), 833–860.
- FEHR, E., KIRCHSTEIGER, G. and RIEDL, A. (1993). Does fairness prevent market clearing? *Quarterly Journal of Economics*, **108**, 437–459.
- (1996). Involuntary unemployment and non-compensating wage differentials in an experimental labour market. *Economic Journal*, **106**, 106–121.
- FEHR, E., KLEIN, A. and SCHMIDT, K. M. (2001). Fairness, incentives and contractual incompleteness. *Working Paper 72*, Institute for Empirical Research in Economics, University of Zurich.
- FEHR, E. and SCHMIDT, K. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, **114**, 817–868.
- FISCHBACHER, U. (1999). z-tree: Zurich toolbox for readymade economic experiments: Experimenter's manual. *Working Paper, 21*. Institute for Empirical Research in Economics at the University of Zurich.
- GNEEZY, U. (2004). Do high wages lead to high profits? An experimental study of reciprocity using real effort. *Working paper*, University of Chicago Graduate School of Business.
- GORDON, D. M. (1990). Who Bosses Whom? The Intensity of Supervision and the Discipline of Labor. *The American Economic Review*, **80**(2), 28–32.
- GROSHEN, E. L. and KRUEGER, A. B. (1990). The Structure of Supervision and Pay in Hospitals. *Industrial and Labor Relations Review*, **43**(3), 134–146.
- HANNAN, R. L., KAGEL, J. H. and MOSER, D. V. (2002). Partial gift exchange in an experimental labor market: Impact of subject population differences, productivity differences and effort requests on behavior. *Journal of Labor Economics*, **20**(3), 923–951.
- KESER, C. and WILLINGER, M. (2000). Principals' principles when agents' actions are hidden. *International Journal of Industrial Organization*, **18**, 163–185.

- KIRCHLER, E., FEHR, E. and EVANS, R. (1996). Social exchange in the labor market: Reciprocity and trust versus egoistic money maximization. *Journal of Economic Psychology*, **17**, 313–341.
- KRUSE, D. (1992). Supervision, Working Conditions, and the Employer Size-Wage Effect. *Industrial Relations*, **31**(2), 229–249.
- LEONARD, J. S. (1987). Carrots and Sticks: Pay, Supervision, and Turnover. *Journal of Labor Economics*, **5**(4), 136–152.
- LUFT, J. (1994). Bonus and penalty incentives Contract choice by employees. *Journal of Accounting and Economics*, **18**, 181–206.
- MINKLER, L. (2004). Shirking and Motivation in Firms: Survey Evidence on Worker Attitudes. *International Journal of Industrial Organization*, **22**, 863–884.
- NAGIN, D. S., REBITZER, J. B., SANDERS, S. and TAYLOR, L. J. (2002). Monitoring, Motivation, and Management: The Determinants of Opportunistic Behavior in a Field Experiment. *American Economic Review*, **92**(4), 850–873.
- NEAL, D. (1993). Supervision and Wages Across Industries. *The Review of Economics and Statistics*, **75**, 409–417.
- PRENDERGAST, C. (1999). The provision of incentives in firms. *Journal of Economic Literature*, **37**(1), 7–63.
- RABIN, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, **83**, 1281–1302.
- (2000). Risk aversion and expected-utility theory: A calibration theorem. *Econometrica*, **68**(5), 1281–1292.
- REBITZER, J. B. (1995). Is there a trade-off between supervision and wages? An empirical test of efficiency wage theory. *Journal of Economic Behavior and Organization*, **28**, 107–129.