

Dittrich, Dennis A. V.; Kocher, Martin G.

Conference Paper

Monitoring and Pay: An Experiment on Contract Design with Endogenous Monitoring

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik
- Session: Incentive Systems, No. D04-V1

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Dittrich, Dennis A. V.; Kocher, Martin G. (2014) : Monitoring and Pay: An Experiment on Contract Design with Endogenous Monitoring, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik - Session: Incentive Systems, No. D04-V1, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:

<https://hdl.handle.net/10419/100580>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Monitoring and Pay: An Experiment on Contract Design with Endogenous Monitoring

Dennis A.V. Dittrich^a, Martin G. Kocher^b

^a*School of Humanities and Social Sciences, Jacobs University Bremen, Campus Ring 1,
D-28759 Bremen, Germany*

^b*Department of Economics, University of Munich, Geschwister-Scholl-Platz 1, D-80539
Munich, Germany*

Abstract

We implement a laboratory experiment in which a principal has to decide on monitoring intensity and pay to investigate whether they are complements or substitutes. Wage level, monitoring intensity, and consequently the desired enforceable effort level are jointly determined by the maximization problem of the firm. As a result, theoretically monitoring and pay should be complements. The between-treatment variation in our experiment is qualitatively in line with the normative prediction of the model under standard assumptions. Also the subject-specific variation indicates that monitoring and pay are considered as complements. Yet, we also find evidence for reciprocal behavior. Our data analysis shows, however, that it does not pay for the principal to only rely on the reciprocity of his agents.

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

JEL-Classification: C91, J31, J41

Email addresses: davd@economicscience.net (Dennis A.V. Dittrich),
martin.kocher@lrz.uni-muenchen.de (Martin G. Kocher)

1. Introduction

While incomplete contracts between principals and agents have been studied extensively in economics, the connection between payments and monitoring in such relationships has only attracted limited attention. Given the importance of an optimal mix of direct monetary incentives (like wages in an employer-employee relation) and monitoring, the scarcity of empirical work is surprising.¹ 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 wages. 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.²

In this paper we will rely on laboratory experiments to empirically assess the design of 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 question whether wages and monitoring are indeed used as substitutes as intuition predicts or rather as complements. For this end, we put four treatments with different parameterizations of productivity and monitoring costs to an experimental test. These four settings can be viewed as representing a range of different industries or firms 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 (2007), and Fehr and Gächter (2002). Two distinctive features of our approach render our setup, however, more general: (i) we implement, as far as we know for the first time, truly continuous effort (costs) for the agent, and (ii) in contrast to earlier studies we apply an endogenous monitoring technology. Therefore, we believe that our experimental test on contractual design and the effects of reciprocity is the most general, so far.

It were Allgulin and Ellingsen (2002), who challenged the conventional wisdom of the substitutive relation between wages and monitoring. Their theoret-

¹In their extensive survey, List and Rasul (2011) report only on one related field experiment. All the laboratory studies on monitoring and shirking surveyed by Charness and Kuhn (2011) focus rather on motivational crowding out driven by negative reciprocity and none on the relationship of payments and monitoring.

²The 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".

ical model shows that monitoring and the level of pay are negatively related, i.e. substitutes, only under two 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, at least as long as common knowledge of rationality and selfishness is assumed.

However, to what extent their results are indicative for actual behavior of subjects in principal-agent relations involving both a compensation and a costly monitoring option is an open question. In reality, behavioral regularities such as 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 (among the first experiments is Fehr, Kirchsteiger, and Riedl (1993); for a recent field experiment see, e.g., Bellemare and Shearer (2009)). The existence of social preferences might completely overturn the theoretical conclusion with regard to the complementarity of monitoring and pay. It is, for instance, straightforward to show that a sufficient number of reciprocal agents in the population renders any monitoring simply unnecessary, leading principals to offer contracts in equilibrium that do not involve any monitoring.

Therefore, it is ultimately a matter of empirical analysis to substantiate whether monitoring and pay are actually complements or rather substitutes (see also, e.g., Chang and Lai, 1999; Demougin and Fluet, 2001). The most desirable approach to answer this question would involve the analysis of field data on contracts and effort exertion. 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 instruments relative to those that do not value such high effort." Using an experiment is, thus, a remedy and can be viewed as a valuable complement to existing empirical studies.

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, between treatment variation is qualitatively in line with the theory. Additionally, analyzing the subject specific variation, we find a clear and highly significant positive correlation between monitoring and pay. Thus, the predictions of the shirking model are confirmed at two different levels: monitoring and pay are indeed used as complements.

Yet, we also observe 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 optimal strategy for principals in our one-shot setting without an option to build reputation. 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 wage and monitor less.

A striking feature of the data is that, although the reciprocity hypothesis is confirmed, contracts without monitoring fare particularly badly, and this is not a consequence of self-selection of different types of agents into contract offers with and without monitoring. A low intensity of monitoring, however, goes a long way, i.e. strengthens reciprocity a lot.

Another finding in line with the existing literature is that the degree of reciprocity is highly heterogeneous across subjects. In other words, the wage-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. Section 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. Theory

2.1. 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 dismissed, thereafter. Note that the shirker receives her compensation for that last period even if she has not exerted any effort. This assumption is an upshot 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.

In the following we present a simplified static version of the shirking model (analysed by Allgulin and Ellingsen (2002)), in which optimal effort and supervision intensity arise endogenously. We consider a bilateral interaction between a risk-neutral principal and a risk-neutral agent.³ 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 he can invest some money $\mu M(p)$ in a monitoring technology that allows him to verify with an investment-dependent probability

³Assuming risk-neutrality is also okay when deriving specific predictions for our experiment. If experimental participants were not close to risk-neutral over the monetary domain at stake in our experiments, we would have to accept ridiculously high risk-aversion levels for higher stake levels according to calibration results in Rabin (2000).

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) $M'(p) > 0$.

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,

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

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,

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

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

2.2. Taking social preferences into account

Let us extend Allgulin and Ellingsen (2002) and 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 by 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 (see also Loewenstein, Thompson, and Bazerman (1989)). 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 the 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 advantageous inequity.

For the sake of succinctness let us – following Fehr et al. (2007) – 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 who 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. (2007) call a trust contract). A fair agent who accepts 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 relevant 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 equalize the two earning 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 equalize 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. Whether 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 the fraction q of fair people in the population, 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 his profit is smaller than one (i. e. higher wages do not pay off in monetary terms), then he 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 not many 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 measures of job autonomy that aggregate 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 inconclusive. While Sessions (2008) finds an inverse relationship of monitoring and pay in the British 1998 Workplace Employee Relations Survey this relationship diminishes under profit sharing schemes. Groshen and Krueger (1990) also 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 premia 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 premia 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 survey by Minkler (2004) indicates that shirking in firms may not be as much of a problem as suggested by standard economic theory (see, e. g., Eaton and White, 1983; Shapiro and Stiglitz, 1984). 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. (2007), Fehr and Gächter (2002), and based on more formal principal-agent models, Fehr et al. (1996), Keser and Willinger (2000), Anderhub, Gächter, and Königstein (2002), as well as Dickinson and Villeval (2008).

The experimental designs of Fehr et al. (1996, 2007), and Fehr and Gächter (2002) share one peculiar and important design feature: 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 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 an implication of assuming that intentional kindness, i. e. high wages, leads to reciprocally kind reactions, i. e. a high effort choice - one would also expect that choosing a trust contract is perceived as kinder 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, and Dickinson and Villeval (2008) for some evidence showing that too high monitoring intensities may, indeed, crowd out intrinsic motivation.

Finally, there is also a piece of evidence on monitoring and shirking from a controlled field experiment. 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

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

to a reduction in the perceived cost of opportunistic behavior by increased shirking. In contrast, employees with good outside options do not increase shirking by more than other employees when the rate of monitoring declines. This shows that the problem of shirking may indeed not be as problematic as predicted by theory, but it is still prevalent and has to be taken care of by appropriate contract designs.

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. The normative solution of the shirking model assuming selfishness and using these parameters is presented in table 1.

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 shown that for our functions 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

⁴Notice 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 $e \geq 1$.

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 himself against exploitation (e. g., as a consequence of being let-down averse), he will monitor, but also offer a higher wage 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, 2007) at the University of Innsbruck. 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.⁵ There were, however, always two employees more on the labor market than employers.

Upon arrival, students were seated at screened computer terminals divided 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.⁶ Before the first period, subjects were assigned to their roles. They kept their roles throughout the whole experiment. Yet, subjects were completely anonymous and not identifiable, i. e. it was impossible to build reputation for either side of the market.

The sequence of actions within any period was: First, employers simultaneously decided on 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 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. One after another (in random order with the probability distribution being i. i. d. across periods) they could, then,

⁵There were 5 sessions with 14 participants and 3 sessions with 18 participants. The distribution over the four treatments is almost perfectly balanced: Only in the treatment with low productivity and low monitoring costs we had two sessions with 14 participants and none with 18. For every other treatment we had one session with 14 and one session with 18 participants. In our analysis we always tested for a session size effect. The respective variable was never significant.

⁶No student declined to participate due to this requirement. In case of losses they were willing to pay their debts immediately.

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 costs, 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-determined 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 periods would have been less desirable.

After 15 periods the experiment ended, and participants were paid in private. A typical session took less than 90 minutes. Note again that reputation building across periods was impossible, i.e. we implemented repetition in order to account for possible experience effects but the experiments actually captures one-shot interactions.

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 approximately the same in all treatments while off equilibrium earning opportunities differed substantially. Average earnings in the experiment were € 8.64.

6. Experimental results

Let us first discuss contract offers. 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 offers and monitoring probability by applying the seemingly unrelated regression (SUR) approach, i.e. we allow the residual error terms of the wage and monitoring equations to be correlated. To account for the dependency of the data due to repeated measurement we apply the multilevel approach, i.e. we include error terms at the matching group level, the subject

level, and the (residual) observation level. For comparison we also provide separate OLS regressions with cluster-robust standard errors and separate multilevel regressions.⁷ The results are presented in table 2.

On average, participants reacted to the treatment variations in the expected way. All main treatment effects have the right sign. As expected, the effect of productivity on wages and monitoring probability is stronger than the corresponding effect of monitoring costs. While productivity is significant in both, the wage and the monitoring probability equation, monitoring costs is only significant in the monitoring probability equation. The interaction effect of productivity and monitoring costs is insignificant in both equations. Note that all three reported models show very similar treatment effects. Only the multilevel seemingly unrelated regression model shows a larger deviation from the point estimates of the other two models as more of the data's variance is absorbed into the random effects. In the multilevel SUR model, especially the time trends seem to be more subject-specific than systematic. The substantial positive correlation of the residuals between the wage offer and the monitoring probability equations is a first indication that wages and monitoring are used as complements in contract design at the individual level. We will investigate this in more detail later.

On the whole, contract choices are adjusted in the right direction, i.e. high productivity leads to higher wages and a higher monitoring probability, and low monitoring cost leads to more intense monitoring and higher wages. Although the impact of monitoring costs on wages is statistically not significant, it is in the direction predicted by the model. However, wages are too high compared to the normative solution under the assumption of selfishness in all treatments with the exception of HPLM, in which average 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. 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.

Observation 1. *The between-treatment variation is qualitatively in line with the normative prediction of the model: Monitoring and pay are complements as a result of variation in (i) monitoring costs and (ii) productivity.*

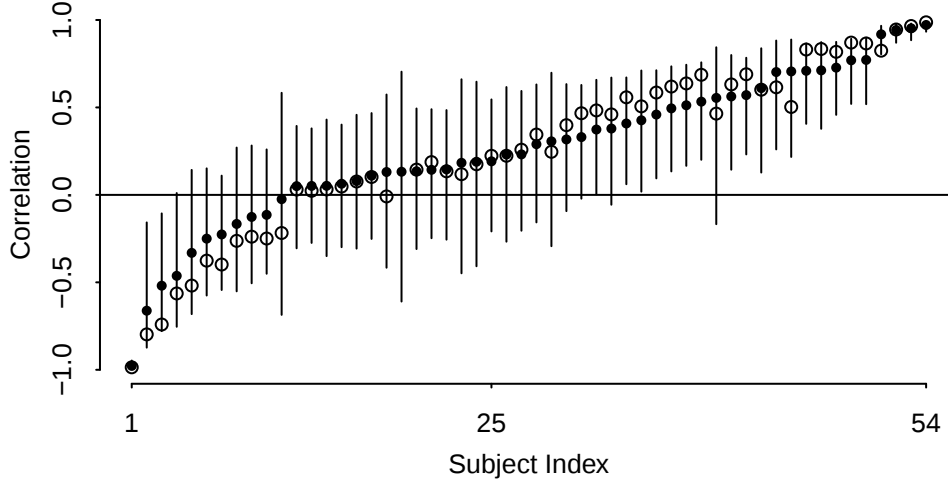
Having established that the between treatment variation in contracts is qualitatively as predicted by the model, we need to verify whether the subject-specific variation is also in line with the normative prediction. Therefore, we compute the correlations between monitoring probabilities and wage offers at the level of the individual principal. Since we cannot assume that individual correlations are independent within a session we, again, estimate the correlations in a multilevel model that allows for partial pooling and may thus improve the efficiency of the estimates while accounting for the possible dependencies. The multilevel model includes a grand mean for the correlations, treatment means that are distributed

⁷If in a set of linear equations all equations shared the same predictors the SUR approach would result in the same estimates as obtained by separate OLS regressions. The multilevel approach to account for the dependency structure of the data, however, introduces random predictors that are not necessarily the same across all equations. Consequently, the SUR approach may increase the efficiency of the estimates.

Table 2: Estimation of wage offer and monitoring probability

Wage Equation	OLS	Multilevel	Multilevel-SUR
Intercept	19.519** (3.883)	19.575** (3.413)	20.032** (1.897)
High Productivity	14.729** (4.491)	14.495** (4.799)	14.389** (2.968)
Low Monitoring Cost	1.509 (4.024)	0.961 (4.890)	0.447 (2.967)
Period	-1.180** (0.141)	-1.180** (0.292)	-0.689** (0.119)
High Productivity : Low Monitoring Cost	-4.942 (4.752)	-3.700 (6.764)	-4.304 (3.599)
High Productivity : Period	0.721** (0.174)	0.721* (0.341)	0.417 (2.948)
Low Monitoring Cost : Period	-0.070 (0.174)	-0.070 (0.341)	0.387 (0.320)
Random Effects Std. Dev. Subjects: Intercept		9.061	8.925
Random Effects Std. Dev. Subjects: Period		1.146	1.177
Correlation Random Effects Intercept - Period		0.340	0.317
Random Effects Std. Dev. Matching Group: Intercept		3.327	3.444
Residual Std. Error	13.280	8.409	8.442
Monitoring Probability Equation			
Intercept	0.341** (0.011)	0.337** (0.052)	0.343** (0.066)
High Productivity	0.218** (0.016)	0.227** (0.073)	0.217* (0.108)
Low Monitoring Cost	0.209** (0.020)	0.219** (0.075)	0.206* (0.109)
Period	-0.018** (0.006)	-0.018** (0.007)	-0.010 (0.019)
High Productivity : Low Monitoring Cost	-0.156 (0.102)	-0.156 (0.111)	-0.133 (0.147)
High Productivity : Period	0.014* (0.006)	0.014* (0.008)	0.006 (0.029)
Low Monitoring Cost : Period	0.013* (0.006)	0.013* (0.008)	0.005 (0.030)
Random Effects Std. Dev. Subjects: Intercept		0.177	0.229
Random Effects Std. Dev. Subjects: Period		0.027	0.142
Correlation Random Effects Intercept - Period		0.395	0.005
Random Effects Std. Dev. Matching Group: Intercept		0.027	0.068
Residual Std. Error	0.289	0.202	0.202
Correlation Residual Errors Wage Eq. - Monitoring Prob. Eq.			0.323

Note: Numbers in brackets are standard errors. For the OLS estimates, standard errors are robust to heteroskedasticity and correlation of arbitrary form within matching groups (HC3 with clusters; see, e.g., MacKinnon and White, 1985). One and two stars indicate significance at the 5 and 1% level respectively. Significance of coefficients of the multilevel models is derived from their simulated posterior distribution. The OLS regressions have an adjusted R^2 of $R^2_{\text{wage}} = 0.23$ and $R^2_{\text{monitoring}} = 0.14$.



Open circles represent individual raw correlations, closed circles are the correlations estimated from a multilevel model, line segments cover the respective 95% credible intervals.

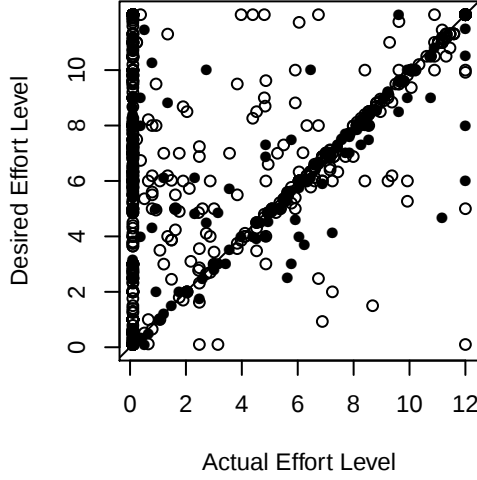
Figure 1: Distribution of individual correlations between monitoring probability and wage offers

around the grand mean, session means that are distributed around the respective treatment means, and individual correlations that are distributed around the session means. Monitoring probabilities and wages were rescaled before estimation of the correlations, i.e. we subtracted the subject-level means and divided by the subject-level standard deviation; any time trends are partialled out.

Figure 1, depicting the distribution of individual correlations between monitoring probabilities and wage offers, reveals a rather large heterogeneity. Individual estimates range from almost perfect negative to almost perfect positive correlations. While only 3 subjects show a significantly negative correlation, 43 subjects show a positive correlation, with 20 being significantly positive. The overall average individual correlation is 0.26 (CI95% = (0.18, 0.33)). There are small differences between treatments. In LPHM average correlation is highest with 0.43 (CI95% = (0.31, 0.54)), in LPLM and HPHM it is 0.28 (CI95% = (0.11, 0.43), (0.19, 0.33)), and in HPLM it is lowest with 0.07 (CI95% = (-0.06, 0.33), $p = 0.09$). It seems that the more favorable the treatment is for the principal, the weaker is the complementarity of monitoring and wages in his contract offers.

Observation 2. *The majority of principals uses monitoring and pay as complements when designing their contracts.*

Wage offer and monitoring probability jointly determine the enforceable effort level (see equation (6)). If the desired effort level $\tilde{e}$ 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 2 presents a scatter-plot of desired 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



Closed circles depict contracts with enforceable desired effort, open circles depict contracts with non-enforceable desired effort.

Figure 2: Desired and actual effort levels

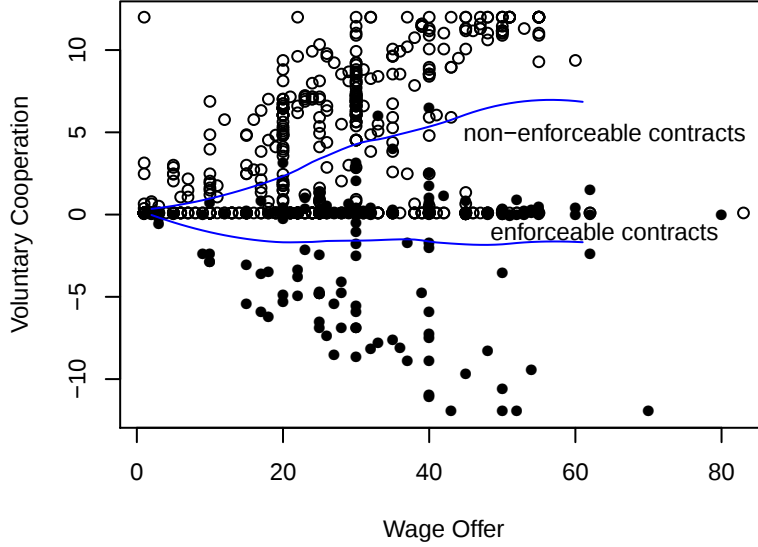
Table 3: Frequency of shirking decisions

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

the desired effort level is enforceable. Second, there is some heterogeneity with regard to effort exertion. While full shirking and exerting just the desired effort level can be easily explained, other decisions either question the conditional rationality of employees or are the result of reciprocity. Yet, despite of the heterogeneity, it is easy to discern from figure 2 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 3 for the frequencies), where we control for repeated measurement and the offered contract, we observe that there is less shirking in the high productivity treatments (likelihood ratio test; $p = 0.01$) and – as one should expect – when the desired effort level is enforceable ($p < 0.01$). 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 ($p < 0.01$) and the offered rent ($p < 0.01$), i.e. wage minus cost of desired effort. This is another piece of evidence corroborating the relevance of reciprocity in our experiment.

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. The latter equals 0 if the contract specifies a non-enforceable desired effort level and equals the desired effort level $\bar{e}$ if the con-



Note: Open circles depict the amount of voluntary cooperation under non-enforceable contracts while closed circles depict 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 3: Voluntary cooperation under enforceable and non-enforceable contracts

tract specifies an enforceable effort level according to equation (6). In figure 3 we show all data points, distinguishing between enforceable and non-enforceable contracts. With non-enforceable contracts the amount of voluntary cooperation cannot be negative. Although the data reveals some 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, however, observe that several agents shirk even under enforceable contracts (in 72 out of 199 cases).

To analyse reciprocity more formally we run a multilevel regression on the amount of voluntary cooperation. The regression model includes random intercepts for the matching group and the individual subjects, treatment dummies, a dummy for zero monitoring, a dummy for enforceable contracts, the wage, different random wage effects on the subjects level for contracts with and without monitoring, the monitoring probability, the desired effort, and the interaction between wage and zero monitoring as well as enforceable contracts and the interaction between monitoring probability and the dummy for enforceable contracts. The multilevel regression results are shown in table 4 together with the results of a corresponding OLS model.

First, we observe no significant 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. This is in contrast to Gneezy (2004) and Dittrich and Ziegelmeyer (2006) who investigate the impact of different productivities on the amount of voluntary cooperation in gift exchange experiments; though it is in

Table 4: OLS and multilevel regression on the amount of voluntary cooperation

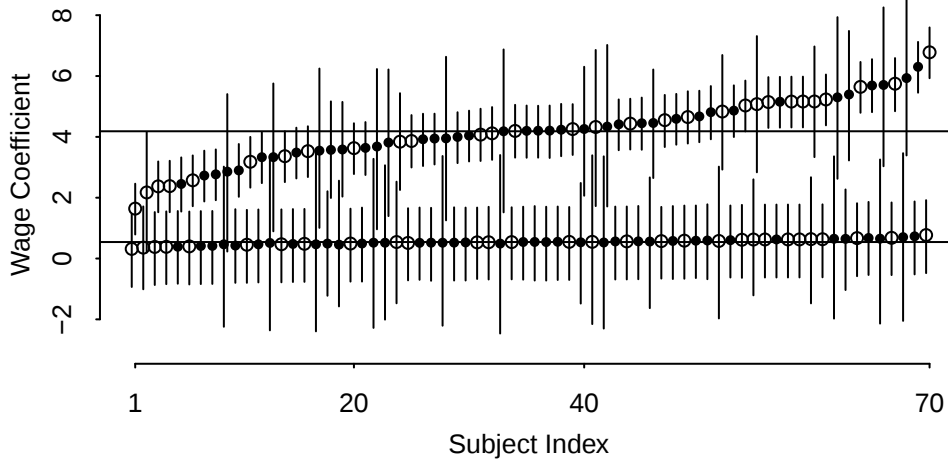
Coefficient	OLS		Multilevel	
	Estimate	Std.Error	Estimate	Std.Error
Intercept	3.192**	0.150	3.246**	0.390
High Productivity	0.745**	0.255	0.829	0.427
High Monitoring Cost	0.592	0.335	0.474	0.421
No Monitoring	-1.927**	0.354	-1.718**	0.584
Enforceable Contract	-5.432**	0.421	-5.559**	0.374
Wage	4.911**	0.809	4.200**	0.512
Monitoring Probability	1.804**	0.414	1.999**	0.428
Desired Effort	-1.684**	0.470	-1.649**	0.445
Period	-0.100*	0.046	-0.105**	0.027
Wage: No Monitoring	-4.128**	1.007	-3.656**	0.752
Wage: Enforceable Contract	-4.519**	0.917	-3.698**	0.669
Monitoring Prob: Enforceable Contract	-2.420*	1.075	-2.802**	0.755
Random Effects Std. Dev. Subjects: Intercept				1.437
Random Effects Std. Dev. Subjects: Wage, $p > 0$				1.781
Random Effects Std. Dev. Subjects: Wage, $p = 0$				0.158
Random Effects Std. Dev. Matching Groups: Intercept				0.001
Residual Std. Error		3.419		3.006

For the OLS estimates, standard errors are robust to heteroskedasticity and correlation of arbitrary form within matching groups (HC3 with clusters; see, e.g., MacKinnon and White, 1985). One and two stars indicate significance at the 5 and 1% level respectively. Period is centered; all other continuous variables are standardized, i.e. raw values are centered and divided by 2 standard deviations. The OLS regression has an adjusted R^2 of 0.43.

line with the observations of Hannan, Kagel, and Moser (2002).

Second, there is a significant positive impact of wage on the amount of voluntary cooperation when the principal monitors. A wage increase of one standard deviation (15.2) leads to a voluntary increase in the effort level of about 2.1 units under monitoring contracts – or a one unit wage increase leads to a 0.14 unit increase in voluntary cooperation. The standard deviation of the random coefficient for the wage offer reveals that participants react rather heterogeneously. Individual wage coefficients vary between 1.64 and 6.78 (cf. figure 4). Yet, if the principal does not monitor agents show no significant reaction to the wage, and there is almost no heterogeneity in their reaction. Individual wage coefficients under no monitoring range from 0.32 to 0.77 with none of these coefficients being significantly different from 0. In order to induce non-minimal effort choices principals have to set a strictly positive monitoring probability. Indeed, under non-enforceable contracts the amount of voluntary cooperation is increasing in the monitoring probability. A monitoring probability increase of one standard deviation (0.31) leads to a voluntary increase in the effort level of about 1 unit. This effect is more than offset under enforceable contracts.

Third, the positive effect of high wages is offset when the contract is enforceable. Under this condition, there is no significant wage–voluntary cooperation relation. Due to some shirking we observe, however, some negative voluntary cooperation independent of the wage. Note that the amount of voluntary coop-



The upper series depicts the wage coefficients under monitoring, the lower series the wage coefficient under no monitoring, i.e. a monitoring probability of $p = 0$. Closed circles represent subjects who never accepted a contract without monitoring, open circles represent subjects who accepted at least once a contract without monitoring. Line segments depict the respective 95% credible intervals.

Figure 4: Distribution of individual wage coefficients in the regression on voluntary cooperation

eration 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, the desired effort level has a negative impact on voluntary cooperation. A one standard deviation (3.55) increase in the desired effort level leads to a decrease of 0.82 units of voluntary cooperation.

The observation that under no monitoring there is also no voluntary cooperation–wage correlation could be due to self-selection of types of subjects into types of contracts. If mainly non-reciprocal participants choose contracts without monitoring, this would explain the observed reduction in the wage–voluntary cooperation correlation. However, our data do not support such a selection effect. 57% of participants in the experiment choose a contract without monitoring at least once. No participant chooses a contract without monitoring more than five times. As figure 4 reveals, there are no systematic differences in the individual wage coefficients in the regression on voluntary cooperation between subjects that always choose contracts with some monitoring and subjects that at least once choose a contract without monitoring. The average individual wage coefficient for the 30 subjects that always choose to be monitored is 4.20. For the 40 subjects that choose not to be monitored at least once it is 4.17. This small difference is statistically not significant (t-test, $p = 0.90$). Both groups are characterized by the same distribution of wage–voluntary cooperation coefficients (Kolmogorov-Smirnov-test, $p = 0.68$). We can, therefore, conclude that to induce non-minimal effort choices, principals have to set a positive monitoring probability, and that this result of our experiment is not due to self-selection of player types into types of contracts.

Observation 3. *If agents are not monitored they do not exert any effort, i.e.*

they maximize their own earnings. In contrast, if they are monitored they behave reciprocally, i.e. exerting more effort than necessary and thus they are forfeiting some earnings. There is a considerable amount of heterogeneity in their degree of reciprocity.

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 alone does not pay. Given the same wage, average per-round earnings are 5.6 (non-enforceable desired effort, $p = 0.02$) to 6.0 points (enforceable desired effort, $p = 0.03$) higher when the employer decides to monitor. This explains why incentive contracts are predominant. In fact, in a multilevel regression on the principal's profit similar to that reported earlier for voluntary contributions the dummy for no monitoring is negative and highly significant while the coefficient on monitoring probability is positive though not significant. A small non-zero monitoring probability already goes a long way. Of course, jointly increasing wage offer and monitoring probability boosts profits further.

Observation 4. *It does not pay for the principal to only rely on the agents' reciprocity. To maximize profits principals should always monitor.*

7. Discussion

We set out on testing whether monitoring and pay are complements or substitutes. In 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, theoretically monitoring and pay should be complements. The main 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, 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 were 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 simplification would, thus, alter the predictions of the model and lead to potentially misleading implications for contractual design.

Although observed contracts are not precisely as predicted by the normative solution of the shirking model 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 wages than implied by the standard model. Analyzing the

subject specific variation, we find a clear and highly significant positive correlation between monitoring and pay. Thus, monitoring and pay are complements. One intuitive behavioral 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 imply that supervision and wages, actually, become complements.

Nevertheless, we also have 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 wages and a higher wage-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 wages, and the amount of voluntary cooperation significantly increases with a higher wage. On the other hand, the latter is not true when there is zero monitoring. Although this last observation is not easy to explain in the framework of common fairness models, reduced levels of reciprocity were also observed by Fehr et al. (1996) and Fehr and Gächter (2002) in their endogenous contract choice treatments. 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 our experimental test, and this result is not a consequence of self-selection of types of players into types of contracts.

Accordingly, at least for the parameters and functions chosen in our experiment relying on reciprocity alone does not pay off. The employer’s earnings are at least as high when he designed a contract with enforceable effort as when he invested the same amount of money in the labor relationship but offered a higher wage and monitors less. This result is somewhat in contrast to recent findings by Falk and Kosfeld (2006), who report that monitoring bears a hidden cost because agents react negatively to being controlled. In their design, however, the principal can only restrict the choice set of agents (this restriction captures the possibility of guaranteeing a minimal effort exertion through control) but not concurrently increase wages. Hence, supervision is always bound to be perceived as distrust in the framework of Falk and Kosfeld (2006), which does not have to be the case in our more general setup. Obviously, the perception of decisions of principals play a crucial role in determining the level of reciprocity, and even subtle design issues might influence this perception and, consequently, behaviorally optimal contractual design.

Finally, as already Kirchler, Fehr, and Evans (1996) observed, there is considerable heterogeneity among agents. The wage-effort relation is by no means the same for everyone. Even though the correlation is on average significantly positive, there is a substantial number of participants who do not or almost not react to the wage, while there are also participants who react very strongly to changes in the offered wage. A similar degree of heterogeneity can be found among the principals: the individual correlations between monitoring and pay differ substantially.

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 (for the parameters that we

use in our experiment and for a one-shot interaction). The four treatment conditions that we put to a test span a relatively large space of feasible parameter combinations. Note that other experiments have shown that the positive effects of reciprocity are strongest for specific environments in which, for instance, high productivities and complete certainty on payoff consequences of one's actions prevail. As these conditions are not (always) met in the real world, it seems important to use different setups to learn more about the nature of reciprocity on labor markets.

One final implication of our findings is that in relations between an employer and an employee, already a low-intense incentive seems to go a long way. Even if it would be optimal to establish strong monitoring mechanisms (e.g., when they are almost costless), it might be a good idea to cut down monitoring to a lower intensity and provide some contractual freedom. However, full trust and zero monitoring does not seem to induce high enough levels of reciprocity to make it profitable, on average.

Appendix A. Appendix - Experimental instructions

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.

Appendix A.1. Welcome

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.

Appendix A.1.1. 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.

Appendix A.1.2. 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.

Appendix A.1.3. 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.

Appendix A.1.4. 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.

Appendix A.1.5. 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.

Periode	1 von 15	Verbleibende Zeit [sec]: 86
Sie haben die Rolle des Arbeitgebers. Wählen Sie bitte die Bedingungen Ihres Vertrages.		
Hier können Sie berechnen, welche Kosten Ihre Kontrollwahrscheinlichkeit für Sie verursachen würde. Kontrollwahrscheinlichkeit: 0.0 / _____, 1		
		Berechnen
Hier können Sie berechnen, welchen Benefit ein gewünschter Effort erzeugt und welche Kosten er für den Arbeiter verursachen würde. gewünschter Effort: 0.1 / _____, 12		
		Berechnen
Kontrollwahrscheinlichkeit:		
Lohn:		
Gewünschter Effort:		
		OK

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

OK

Kein Vertrag

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:

Periode 1 von 15 Verbleibende Zeit [sec]: 26

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 / 1.2

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

Berechnen

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.

Appendix A.1.6. How profits are calculated

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 = $16 \cdot (\text{actual effort})^{(2/3)} - 20 \cdot (\text{monitoring probability})^2 - \text{wage}$

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

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.

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 = wage - (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.

Appendix A.1.7. 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., Ellingsen, T., 2002. Monitoring and Pay. *Journal of Labor Economics* 20 (2), 201–216.
- Anderhub, V., Gächter, S., 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.
- Arai, M., 1994b. An Empirical Analysis of Wage Dispersion and Efficiency Wages. *Scandinavian Journal of Economics* 96, 31–50.
- Athey, S., Stern, S., 1998. An Empirical Framework for Testing Theories about Complementarity in Organizational Design. Working Paper 6600, National Bureau of Economic Research.
- Bellemare, C., Shearer, B., 2009. Gift giving and worker productivity: Evidence from a firm-level experiment. *Games and Economic Behavior* 67, 233–244.
- Bénabou, R., Tirole, J., 2003. Intrinsic and extrinsic motivation. *Review of Economic Studies* 70, 489–520.
- Bolton, G., Ockenfels, A., 2000. A theory of equity, reciprocity and competition. *American Economic Review* 100, 166–193.
- Chang, J., Lai, C., 1999. Carrots or sticks? A social custom viewpoint on worker effort. *European Journal of Political Economy* 15, 297–310.
- Charness, G., Kuhn, P., 2011. Lab labor: What can labor economists learn from the lab? In: Ashenfelter, O., Card, D. (Eds.), *Handbook of Labor Economics*. Vol. 4A. Elsevier, Amsterdam, Ch. 3, pp. 229–330.
- Demougin, D., Fluet, C., 2001. Monitoring versus incentives. *European Economic Review* 45, 1741–1764.
- Dickinson, D., Villeval, M. C., 2008. Does monitoring decrease work effort?: The complementarity between agency and crowding-out theories. *Games and Economic Behavior* 63, 56–76.
- Dittrich, D. A. V., Ziegelmeyer, A., 2006. Laboratory bilateral gift exchange: The impact of loss aversion. *Papers on strategic interaction*, Max Planck Institute of Economics, Jena.

- Eaton, C., White, W. D., 1983. The Economy of High Wages: An Agency Problem. *Economica* 50, 175–181.
- Falk, A., Kosfeld, M., 2006. The hidden cost of control. *American Economic Review* 96, 1611–1630.
- Fehr, E., Gächter, S., April 2002. Do incentive contracts crowd out voluntary cooperation. Working Paper 34, Institute for Empirical Research in Economics, University of Zurich.
- Fehr, E., Gächter, S., Kirchsteiger, G., 1997. Reciprocity as a contract enforcement device: Experimental evidence. *Econometrica* 65, 833–860.
- Fehr, E., Kirchsteiger, G., Riedl, A., 1993. Does fairness prevent market clearing? *Quarterly Journal of Economics* 108, 437–459.
- Fehr, E., Kirchsteiger, G., Riedl, A., 1996. Involuntary unemployment and non-compensating wage differentials in an experimental labour market. *Economic Journal* 106, 106–121.
- Fehr, E., Klein, A., Schmidt, K. M., 2007. Fairness and contract design. *Econometrica* 75, 121–154.
- Fehr, E., Schmidt, K., 1999. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* 114, 817–868.
- Fischbacher, U., 2007. z-tree: Zurich toolbox for readymade economic experiments. *Experimental Economics* 10 (2), 171–178.
- Gneezy, U., April 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., 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., 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., Willinger, M., 2000. Principals’ principles when agents’ actions are hidden. *International Journal of Industrial Organization* 18, 163–185.
- Kirchler, E., Fehr, E., 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.

- List, J. A., Rasul, I., 2011. Field experiments in labor economics. In: Ashenfelter, O., Card, D. (Eds.), *Handbook of Labor Economics*. Vol. 4A. Elsevier, Amsterdam, Ch. 2, pp. 103–1228.
- Loewenstein, G. F., Thompson, L., Bazerman, M., 1989. Social utility and decision making in interpersonal contexts. *Journal of Personality and Social Psychology* 57 (3), 426–441.
- Luft, J., 1994. Bonus and penalty incentives Contract choice by employees. *Journal of Accounting and Economics* 18, 181–206.
- MacKinnon, J. G., White, H., 1985. Some heteroskedasticity consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics* 29, 53–57.
- 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., 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.
- Rabin, M., 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.
- Sessions, J. G., 2008. Wages, supervision and sharing. *Quarterly Review of Economics and Finance* 48 (4), 653–672.
- Shapiro, C., Stiglitz, J. E., 1984. Equilibrium unemployment as a worker discipline device. *American Economic Review* 74, 433–444.