Pay – But Don't Pay Too Much
An Experimental Study on the Impact of Incentives[*]

Kathrin Manthei[**]

University of Cologne

Abstract: Most Principal-Agent models predict that increasing incentives result in higher performance. This paper examines whether this result is valid under real-effort conditions. Exposing the participants to varying strengths of incentives we find an inversely U-shaped relationship between effort levels and incentive intensity, which not only contrasts predictions of standard theory but also observations in previous real effort experiments. We provide a new theoretical explanation for the results within a principal agent model with loss averse agents.

Key words: compensation; incentives; loss aversion; real effort
JEL Classification: M52, J33

[*] Financial support was provided by DFG KR2077/2-1. I thank Dirk Sliwka for helpful comments and assistance.
[**] Universität zu Köln, Seminar für Allgemeine Betriebswirtschaftslehre und Personalwirtschaftslehre, Herbert-Lewin-Str. 2, 50931 Köln, Tel: +49-221-4702584, Fax : +49-221-5078, Email: kathrin.manthei@uni-koeln.de

Pay – But Don't Pay Too Much

An Experimental Study on the Impact of Incentives

1. Introduction

The question of optimal incentive schemes in Principal-Agent relationships with hidden action has been the object of research for many years. A well-established result of most standard hidden action models is that higher incentives ceteris paribus lead to higher performance. Field studies and experiments present evidence for this conclusion (Lazear 2000; Paarsch/Shearer 1999; Dickinson 1999).

Still standard results are controversial. Several experimental studies in economics were recently able to show what psychologists have already been claiming for some time namely that higher incentives do not inevitably stimulate higher effort choices. Moreover, cases exist where introducing an incentive contract even reduces effort. This phenomenon emerges in real effort experiments (Gneezy/Rustichini 2000a; Gneezy/Rustichini 2000b) as well as in experiments including abstract effort choices (e.g. Fehr/Gaechter 2002; Irlenbusch/Sliwka 2003)[1].

A frequently cited key concept explaining certain anomalies is motivation crowding out (e.g. Frey 1997). In this framework two types of motivation are specified, namely intrinsic and extrinsic motivation[2]. The core of the theory in this context is that implementing a performance dependent compensation scheme might strengthen extrinsic incentives while diminishing intrinsic motivation. As a consequence the effect of introducing a variable compensation is ambiguous. Frey (1997) introduces a

[1] In these experiments effort choices are typically represented by abstract numbers, which the participants can choose. With higher numbers the agent's pay-off decreases and the principal's pay-off increases.

[2] Intrinsic motivation denotes an inner drive to do things (e.g. pleasure) while extrinsic motivation describes a behavior driven by rewards outside the individual. The idea of intrinsic motivation was first introduced by Deci (1971) and deCharms (1968).

Principal-Agent model including intrinsic motivation. The agent's utility increases in wage and decreases in effort. If the principal implements stronger incentives the impact of the intervention on the agent's effort choice is not clear. Frey distinguishes three effects caused by the principal's intervention namely a price effect, an enhancing effect and a crowding out effect. The price effect simply denotes the effect of higher opportunity costs of lower effort levels. Representing a positive perception of the intervention the enhancing effect amplifies the impact of the price effect. In turn the crowding out effect refers to a negative assessment of the increase of incentives. The principal's intervention undermining the agent's intrinsic motivation produces lower effort choices. Still the net effect is hardly predictable since price and enhancing effect point in the same direction, while the crowding out effect affects the opposite. Particularly is seems conceivable that for lower interventions the crowding out effect dominates the price effect resulting in reduced effort whereas higher interventions cause increasing effort choices due to the prevailing influence of the price effect[3].

However the concept of intrinsic motivation is not undisputed among psychologists (e.g. Eisenberger/Cameron 1996; Deci/Koestner/Ryan 1999a; Deci/Koestner/Ryan 1999b)[4]. Especially in experiments with abstract effort choices motivation crowding theory fails to provide sensible explanations, as there is no task stimulating intrinsic motivation.

Another part of the related literature indicates the relevance of reference dependent preferences (e.g. Fehr/Goette 2002; Camerer, Babcock, Loewenstein & Thaler 1997). Generally the effect of wage variations on labor supply and effort exertion is being examined. Fehr/Goette (2002) find decreasing effort choices with stronger incentives

[3] This is only one of various theories on the concept of intrinsic motivation.
[4] For a good overview on intrinsic motivation see Kunz/Pfaff (2002)

for reference dependent preferences if the reference income has been exceeded. In that case higher piece rates have a diminishing impact on work effort.

The controversial evidence on the effect of incentives in work relationships illustrates that further investigations of this question are necessary.

This paper investigates the influence of varying piece rates on work performance. For that purpose we conducted two real effort experiments at the Universities of Bonn and Cologne. Real effort conditions were chosen to prevent subjects from restraining on income distribution and to generate noticeable disutility from higher effort. The experimental design of both experiments has been inspired by a real effort experiment conducted by Gneezy/Rustichini (2000a). They investigated the influence of varying incentives on effort in an IQ-Test task. Participants were separated into four groups. One group was paid a participation fee only. The other groups were paid the participation fee plus additional 0.1, 1, 3 NIS for every correct answer in the IQ-Test[5]. The results showing a non-monotonic relation between incentives and scores contradict standard theoretical predictions. The group without any incentives (paid the participation fee only) outperformed the group with additional 0.1 NIS/correct answer. The two other groups (additional 1 and 3 NIS) yielded significantly higher scores than the 0,1 NIS group. So in total Gneezy/Rustichini (2000a) found a V-shaped relationship between effort and intensity of incentives. Referring to motivation crowding theory one could argue that for the low incentive group the crowding out effect dominated the price effect while for the higher incentive treatment the price effect prevailed. As another explanation Gneezy/Rustichini (2000a) offer an explanation based on of incomplete contracts stating that the fixed wage group thought the work on the IQ-Test to be part of the contract concluded with the experimenters. Introducing the variable compensation

[5] NIS=New Isreali Shekel names the Israeli Currency.

completed the contract with regard to meaning of the fixed wage as a show-up fee. According to Gneezy/Rustichini (2000a) the variable component then was perceived as the compensation for working on the task.

Still there were some essential modifications in our experiments. Considering the literature mentioned above it is obvious that the effect of incentives is not clear. So a key aim of the experiments was the investigation of the impact of incentives on individual effort choices and in doing so to check the robustness of the Gneezy/Rustichini (2000a) results. Deviating from Gneezy/Rustichini (2000a) we introduced some essential modifications concerning the compensation schemes. In Gneezy/Rustichini's (2000a) design only about 0.38% of the wage of the low payment treatment was performance-dependent. For the two higher paid groups the fractions were roughly 6% and 15%. The low payment group produced the worst results. A possible explanation for this is that the incentive parameter was too low in comparison to the fixed wage, so that it was not noticeable and hence could not evolve its full incentive impact. In fact we applied a much higher ratio of fixed to variable wage components to ensure a noticeable wage increase with increasing effort and to check the robustness of the V-shaped relationship.

As motivation crowding out seems to be a possible explanation for the failure of incentive contracts the second key objective of the experiments was to examine how far qualities of tasks can influence the subjects' effort. The Type of tasks might be important with regard to their ability of stimulating intrinsic motivation. We test for this by introducing another task in addition to an IQ-Test. Hence we implemented two tasks overall with putative different ability of stimulating intrinsic motivation. The first one corresponded closely to the IQ-Test Gneezy and Rustichini (2000a) used. As a second task subjects were asked to count the frequency of a particular number in blocks of

random numbers. It was different with respect to some features. First it did not consist of various exercises but remained the same during the whole time of the experiment. Further it was monotone, boring and required a lot of concentration.

Since we find the highest effort choices with an intermediary incentive payment the experimental results of the first experiment, conducted at the University of Bonn, do not give evidence for an effect of motivation crowding out. Developing an inversely U-shaped relationship between effort choices and level of incentives the outcome does not confirm standard theory either. Moreover it indicates the influence of reference dependent preferences and loss aversion namely, that individuals adjust their effort choice with respect to a reference income. Therefore we develop a simple Principal-Agent model accounting for the agent's loss aversion. The second experiment conducted at the University of Cologne was run to test the robustness of the results.

The remainder of this paper is structured as follows: Section 2 describes the experimental design and procedure and develops hypotheses on the results. Furthermore the results are presented and interpreted. In section 3 the model including reference dependency is introduced and developed. The last section concludes.

2. The real effort experiments

Experimental Set-up and Procedures

The experiments were conducted at the Universities of Bonn and Cologne in November 2002 and July 2003 respectively. In total 209 of the Universities' undergraduate students of various disciplines participated in the experiments. The students were randomly assigned to groups with different tasks and wages. There were six different treatments with two different tasks and three different levels of variable compensation. For an illustration see Table 1. The first task was comparable with Gneezy/Rustichini's

(2000a) IQ-Test (IQT) consisting of extracts from a book containing exercises for logical training. No special skills were necessary to answer these questions. With the other task (CN) participants were required to count the number of "ones" and "sevens" out of a block of random numbers put together by a computer program[6].

Table 1

Treatments

task/performance-contingent wage	0€	0.05€	0.5€
CN			
IQT			

All participants were told that they would always get 5€ as a participation fee. The participants of the first two treatments received the participation fee only (No Incentive=NI). The other groups earned an additional 0.05€ (low incentive=LI) and 0.5€ (high incentive=HI) respectively for every point scored[7].

The subjects who worked on the IQT received a point for every correct answer but 0.5 points were subtracted for any incorrect answer[8]. In the CN an answer was valued correct if it corresponded to the correct number of "ones" and "sevens" respectively with a deviation of one. If the exact number of "ones" in a block was for example 30, subjects who counted 29, 30 or 31 "ones" received one point. After all individuals had entered the corresponding room they were requested to sit down on seats marked with a pen and pieces of paper[9]. After everybody had sat down the supervisors handed out the

[6] For an example see appendix.
[7] 1€ was about 1$ at the time of the experiments.
[8] That was to prevent subjects from guessing which seemed to be necessary as there were several multiple-choice questions.
[9] In the session at the University of Bonn students were separated according to their payment scheme. Only participants with the same incentive scheme worked in the same room. In contrast to that

exercises. Each of them was covered with a sheet with the instructions printed on[10]. The subjects were given 30 minutes to work on the tasks. After exactly 30 minutes a bell rang and the supervisors handed out forms in which the subjects were asked to fill in their answers. Additionally a questionnaire was fixed to the form[11]. The participants were granted another 5 minutes to copy the answers into the form and answer the questionnaire. The questionnaire included several questions such as gender, age, years of study etc. The supervisors collected the forms and informed the subjects when and where they could come and receive their payment. The whole procedure took about 45 minutes.

Hypotheses

According to standard theoretical predictions participants of treatments with higher incentives exhibit a higher performance. Explicitly we consider higher point scorings to approximate higher effort choices. The participants in the treatments with the fixed wage are expected to exert the lowest effort level. Members of the LI treatments should perform better than members of the NI treatments while participants of the HI treatments exert even more effort and therefore are supposed to yield the best result independent of task. This relation can be expressed by

$$e^{NI} < e^{LI} < e^{HI}.$$

Considering motivation crowding theory things appear different. Due to the multitude of different effects an unambiguous prediction is difficult to make. Assuming that for small interventions the crowding out effect dominates the price effect but for high

participants of all treatments worked in one single room in the session at the University of Cologne. This measure was implemented to eliminate room effects.

[10] A translated version of the instructions is available from the author on request.

[11] Data on subjects who did not fill in the questionnaire were extracted from the sample. That was necessary because they might have continued working on the tasks, which adulterates their scores.

interventions the price effect dominates the crowding out effect participants in the LI treatments should perform worse than those in the NI treatments. Participants in the HI treatments in turn should yield higher scores than those in the NI treatments. As the IQT treatments resemble Gneezy/Rustichini's (2000a) experiment a similar effect could be expected i.e. the treatment with fixed wage only and the high incentive groups exert significantly more effort than the low incentive group. Analogously to that the relation is

$$e^{NI} > e^{LI} < e^{HI}.$$

Due to the monotony of the CN task different observations were likely because compared to the IQT its potential to stimulate intrinsic motivation or work pleasure can be assumed to be lower. Thus a monotonously increasing relation between wage and effort with these treatments (CN) could be expected:

$$e^{NI} < e^{LI} < e^{HI}.$$

Results

Table 2 in the appendix presents the averages of points scored in the different treatments at the University of Bonn and Cologne. As presented in Figure 1 the median number of points in three of four cases show that effort increases from no incentive to low incentive treatments. According to standard theory this outcome is not surprising. Yet it seems to contradict Gneezy/Rustichini's (2000a) findings. They present an exactly opposite result since their fix paid treatment outperformed their low incentive treatment. Furthermore, results decrease in all cases comparing the LI and the HI group, which indicates lower effort choices with the HI treatments. This result cannot be explained by

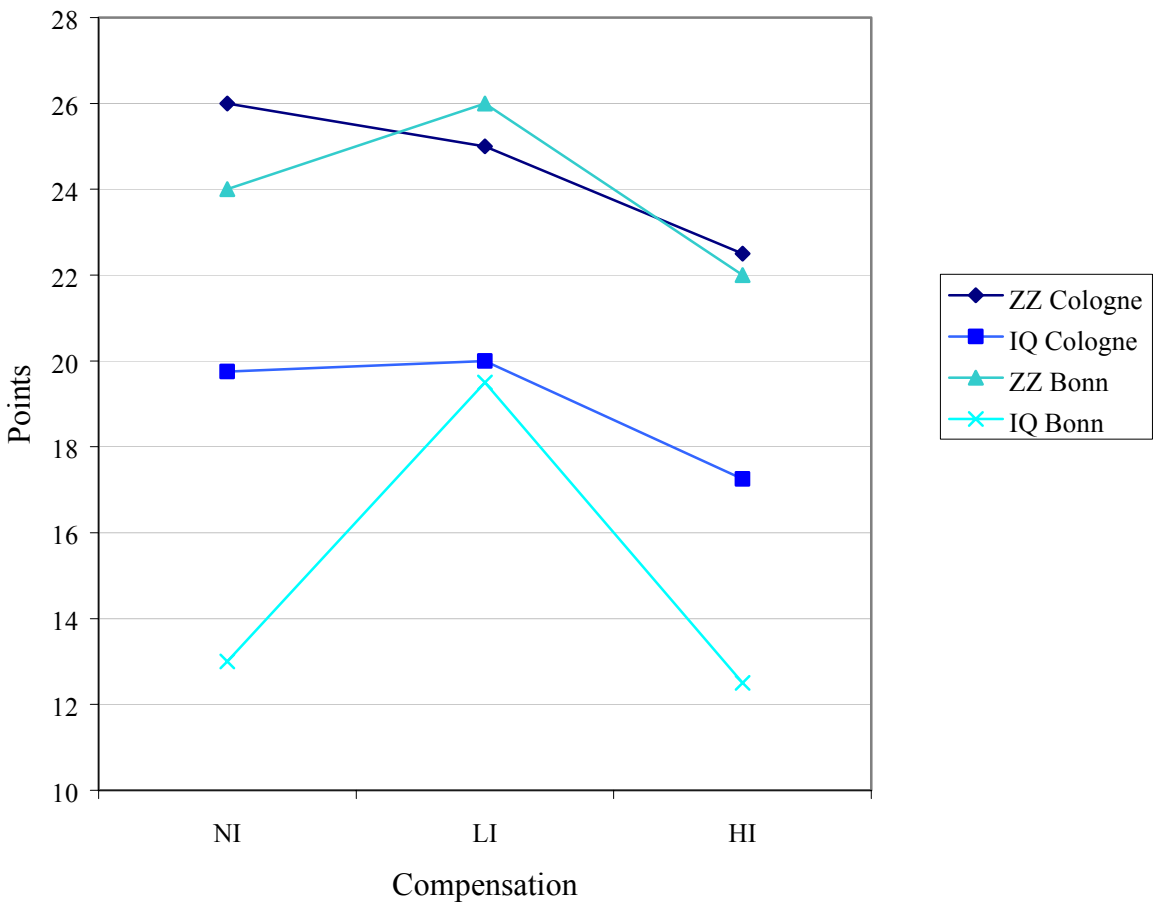

Figure 1

standard theory, as higher incentives should stimulate higher effort choices. Again the result is not compatible with Gneezy/Rustichini's (2000a) findings either since they find a positive relationship between performance and incentives if once introduced. The NI group and the HI group fall behind the LI group in both tasks. That is, subjects in the LI treatment achieve the best results independent of task.

Hence, it is obvious that the results contradict standard theoretical predictions. There is no monotonous increase of effort with the incentive parameter. We neither find an outcome similar to the results by Gneezy/Rustichini (2000a) but observe the opposite pattern

$$e^{NI} < e^{LI} > e^{HI}.$$

It seems to be surprising that the Low Incentive Group outperformed the two other groups. In three of four cases they form a peak to which the other groups fall back.

The data were analyzed with median regression. This seems to be appropriate here as median regression minimizes the sum of absolute deviations instead of the sum of squared deviations. That makes it less sensitive towards outliers and thus is a better measure for the central tendency of the data[12]. As the data show several extreme scores the method fits very well with it[13].

Data of both tasks and both universities were pooled for this analysis and the low incentive scheme was taken as a reference level of compensation. As stated above individual point scorings are used as proxy for exerted effort on the task. Estimates from Median Regression are presented in Table 4.

The descriptive statistics illustrated that in three of four cases both deviations from the reference level of compensation (low incentive) lead to lower performance. And indeed, the high powered incentives lead to highly significant and sizeable negative effect on performance. The median regression shows that individuals in the HI treatments achieve significantly 5,47 points less than those in the LI treatments.

The decrease of effort with a higher rate of compensation seems to be the most interesting discovery within the data. Table 4 shows a negative and highly significant influence of the high incentive scheme on point scorings. In particular we can neither find an increase in scorings with changing from a low performance contingent compensation to a higher one nor a decrease in scoring comparing the fixed wage with a

[12] For example there were subjects who had negative scorings in the IQT.

[13] For an analysis by ordinary least squares see Table 5 in the appendix. A separate analysis of the Bonn and Cologne treatments is presented in Table 3 in the appendix. Investigating the figures of the Bonn treatments by median regression the coefficients for the dummy variables "High" is –5 significant at a 5 percent level of significance while the coefficient for the dummy variable "Fix" (fix wage) is –4 and weakly significant. (See Table 3, first column) Results for the treatments conducted at the University of Cologne are similar but only the coefficient for the "High" Dummy is significant.

Table 4

Models of point scorings
Estimates from median regression
Dependent variable: Points

Pooled Data

Explanatory variable	Regression coefficients (t-statistics)
	Model
Intercept	36.37*
	(1.54)
Fix	-0.97
	(-0.60)
High	-5.47***
	(-3.47)
Cologne	1.89
	(1.44)
IQ	-5.00***
	(-3.80)
Sem	1.74***
	(2.96)
Sem²	-0.17***
	(-3.32)
Age	-0.85
	(-0.45)
Age²	0.01
	(0.27)
No. Of Observations	181
Pseudo R²	0.1165

Fix is a dummy variable indicating that an individual is member of the No Incentive Group. High is a dummy variable with value one if the individual is member of the High Incentive Group. IQ is a dummy variable with value one if the individual worked on the IQ-Test. Cologne is a dummy variable with value one if the individual participated in the session at the University of Cologne. Sem is the number of semesters the individual has already spent at university. Sem² is the squared number of semesters the individual has already spent at university.

small performance dependent compensation. As the results contrast Gneezy/Rustichini's (2000a) findings these experiments do not provide evidence for an undermining effect of incentives of the form described in the hypothesis in section 2.2. This seems to be even more visible considering that qualitatively similar outcomes occur independent of task.

Interpretation

A possible interpretation of the results could be an increasing crowding out effect with increasing strength of incentives. That is, the stronger the extrinsic incentive the more intrinsic motivation is crowded out causing reduced effort choices. However, this theory cannot explain the poor results of the NI group since this group's performance should at least be better than the LI group's. Another explanation might be a growing error rate with increasing incentives. Intuitively one could argue that participants might feel excited imagining a very high income. Therefore the implementation of a high piece rate might stimulate very high effort in quantity resulting in worse quality of performance and consequently more mistakes. In that case a rather high effort in quantity would cause the worse scorings of the HI group. However, the data do not confirm this conjecture. The distribution of given answers (correct and incorrect) is predominantly similar to the distribution of point scorings. There are no significant differences between error rates in different treatments. Therefore this explanation is not very plausible.

To us the most convincing interpretation seems to be a theory of reference dependent preferences meaning that the subjects had a certain reference income in mind when participating in the experiment. At the University of Bonn experiments are run regularly. It is a common habit for students to participate and earn extra money. It is

well known that participation in experiments is remunerated with an average of 10€[14]. On the contrary students were not familiar with taking part in economic experiments at the University of Cologne at the time of the session. Therefore participants had to be recruited from undergraduate courses by reporting the average wage in the experiment of 10€. So in that case it was possible to influence the participants' expectations and generate an artificial reference level. Due to the design of the experiment it was not possible to earn 10€ for the NI group since their payment was independent of their performance. According to standard theory they did not have any incentive to work on the tasks at all. Participants in the LI group could earn a maximum wage of 7.40€ which is less than 10€. Still reaching this maximum income was very unlikely since the average number of points actually reached was much lower than the maximum. So it can be assumed that members of the LI group knew from the start that they would have to work hard and concentrated to get close to their reference wage. In contrast to that the HI group could reach the level of 10€ by attaining few points only. Consequently the LI group had to work a lot harder than the HI group and did so to reach the optimal utility level.

3. A reference wage hypothesis

Standard Principal-Agent results cannot explain the surprising outcomes of the experiments. For that reason a different type of theory has to be established. This section tries to explain the results with a simple model including a reference wage.

The assumption of a reference level of compensation seems to be plausible, since people might not only use their actual wage level to evaluate their utility but take a reference level (e.g. an expected wage or a rival's/workmate's wage) into account.

[14] This is an average value. Of course this value varies within the experiments depending on the role the participant plays. In ordinary jobs students earn about 10 to 15€.

The relevance of reference points has been the object of research in many fields. Among the first, introducing the concept of reference points are Kahneman/Tversky (1979) analyzing decisions under uncertainty. They develop a model describing loss aversion by designing a utility function including a reference point. This utility function has a convex slope below the reference point. It changes abruptly at the reference point and develops an infinite concave slope. Hence subjects behave risk seeking below the reference point and risk averse beyond it. Evidence for the relevance of the theory has early been found in many experiments (e.g. MacGrimmon/Larsson 1979). Investigating the effect of loss aversion under risk-less choice Tversky/Kahneman (1991) extent the application of reference dependent utility. Easterlin (2001) examines the relation between happiness and income. Though he finds a positive relationship between income and happiness, income growth does not affect lifetime happiness since according to his theory aspirations grow with increasing income. Thus, the reference point, from which happiness is evaluated, changes. Fehr/Goette (2002) find reduced effort with higher wages in a field experiment on the work habits of bicycle messengers. The Messengers worked more days a month but decreased their shifts per working day that is reduced effort. In a study on New York Cab Drivers Camerer et. al. (1997) report decreasing numbers of working hours among Cab Drivers on high wage days. The experiments described in this paper show that comparable results can occur in laboratory experiments as well.

In this approach the assumption is made that peoples` utility does not only refer to the absolute height but rather to the relative height of monetary compensations. After the reference level is reached the following growths relatively lose value. It follows that compared to the standard case less or no additional effort is rational after reaching the reference level, as costs would exceed utility gains from wage. The utility function

increases linear in wage but develops a smaller slope as a reference wage is reached. The reference wage is defined as a point from where wages are evaluated. This might be a wage the agent expects or perceives to be appropriate for a certain task. After that point is reached the slope of the utility function flattens. Thus utility increases slower if wages exceed the reference point. Hence the extreme case would be a constant utility level $\bar{v}$. That is, utility is independent of wage beyond the reference point.

The complete model is described in the next section.

The Model

Since the model described in this section is supposed to be a theoretical approach to explain the experimental results we do not calculate the optimal incentive scheme. We rather take a wage contract comparable to those in the experiment. Consequently, the wage contract is an assumption in the model rather than the theoretical result.

Assume a utility function, which is additive-separable of the form:

$$U(w,e)=v(w)-c(e),$$

where w represents wage and e denotes the agent's exerted effort. As shown in Figure 2 the slope of the utility function is discontinuous at value R

$$v(w)=\begin{cases} w & if\ w < R \\ R+(w-R)s & if\ w \geq R \end{cases} \qquad (1)$$

with $s < 1$.

The value R represents the reference wage from which the agent evaluates the actual wage. Since s is less than 1, marginal utility from w is smaller beyond the reference point than below. Assessing the situation from the reference point R the agent is in a loss situation if the first inequality is met because she stays below it. If the second inequality is met she is in a win situation. As the agent is work-averse, effort exertion is

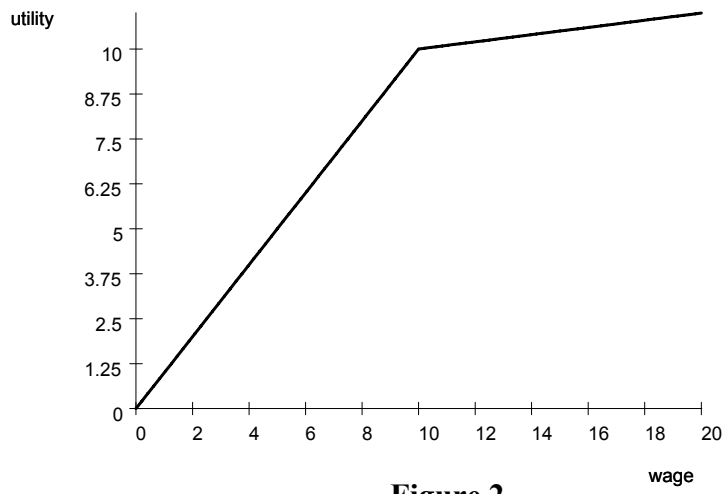

Figure 2

(Plotted for $R = 10, s = 0.1$)

costly. Costs *c(e)* are convex in *e* with $c'(e) > 0$, $c''(e) > 0$, $c(0) = 0$ and $\lim_{e \to \infty} c'(e) = \infty$.

Furthermore let the wage contract be linear in *e* with a fixed wage α:

$$w = \alpha + \beta e .$$

The crucial question is for what value of *e* the agent changes from the left to the right side of the utility function. That is exactly the effort level where the kink is located. Let us call this decisive value the critical $e = e^R$. The employee switches to the right side if *w* at least equals *R*. It follows:

$$e^R = \frac{R - \alpha}{\beta} . \qquad (2)$$

Furthermore the optimal choice of effort is determined by the first order condition of the agent's objective function[15]. The first derivative of the corresponding objective function yields

[15] Due to the concavity of the objective function the first order condition is necessary and sufficient for a maximum.

$$U'(e) = \begin{cases} \beta - c'(e) & \text{if } e < e^R \\ s\beta - c'(e) & \text{if } e \geq e^R \end{cases} \quad .$$

This consideration leads to three possible cases, which will be discussed in the following.

Proposition: *For given values of α and β the agent's optimal effort level e* is:*[16]

$$e^* = \begin{cases} c'^{-1}(\beta) & \text{if } \beta < c'\left(\frac{R-\alpha}{\beta}\right) \\ \frac{R-\alpha}{\beta} & \text{if } \frac{1}{s}c'\left(\frac{R-\alpha}{\beta}\right) \geq \beta \geq c'\left(\frac{R-\alpha}{\beta}\right) \\ c'^{-1}(s\beta) & \text{if } s\beta > c'\left(\frac{R-\alpha}{\beta}\right) \end{cases} \qquad (3)$$

The first case determines the employee's optimal effort choice if it is located on the left hand side of the critical value. The second case presents the effort decision, if it is located exactly on the critical value. The third case shows the employee's effort decision if it is located right to the kink. The change of e^* in the incentive parameter is presented in figure 3.

Figure 3 shows that there is an area, roughly between 2.5 and 7.5, where the optimal effort level chosen by the agent decreases in the incentive parameter. In this area the agent always chooses e^R. Since e^R decreases in β we find a downwards sloping curve in this interval. Consequently from the principal's point of view increasing piece rates in this area are counterproductive not only because they are costly but also because they even reduce the agent's performance.

Evaluating points on the right of the critical value from the reference point the employee experiences gains. Due to the employee's loss aversion these marginal gains are less

[16] For the proof of the proposition see appendix.

profitable and thus more costly than marginal gains on the left side of the reference point.

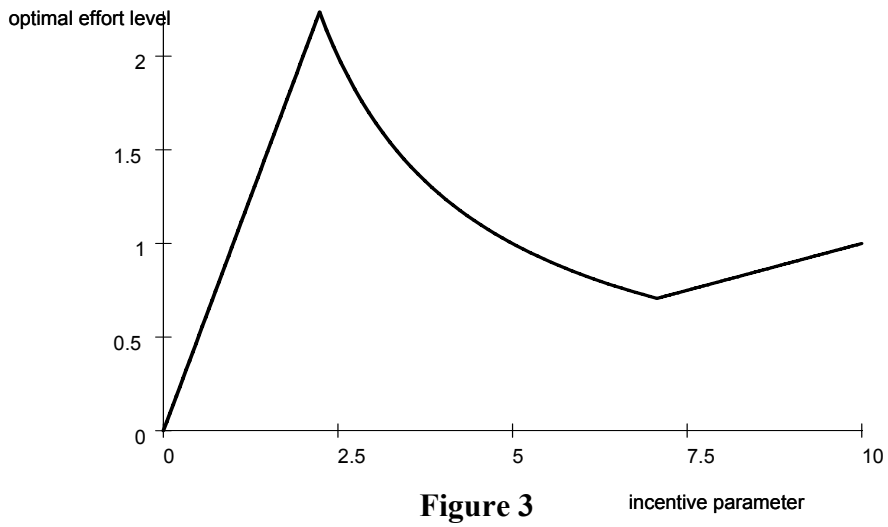

Figure 3

(Plotted for $R = 10, \alpha = 5, c(e) = 0.5e^2, s = 0.1$)

With discrete choices of β, e.g. a high β_H one and a low β_L one as it was in the experiments, it is possible to recreate our experimental results with this model. A necessary condition for this is of course, that the optimal effort choices are located on different sides of the critical effort value.

However the model introduced above leaves room for criticism and development. Naturally the choice of parameters drives the model. Therefore parameters R, β,s and the shape of the cost function decide whether the model generates the experimental results. Unfortunately the predictive power of the model is very low, since obviously a reference wage is the individual's private information and not directly measurable. Furthermore the reference income might vary between individuals or groups of individuals. Assuming different reference wages for different agents would therefore be an interesting modification of the model but exacerbates predictions. Moreover the conditions, under which this model can explain the phenomenon that occurred in the

experiments, are very specific, assuming rather precise wage expectations. Still it can enlighten the special situation in which experiments take place.

4. Concluding Remarks

In this paper the relation between incentives and effort choices has been examined. The experimental results show that effort does not inevitably increase with increasing incentives. Instead the opposite was the case in the experiments conducted, as higher incentives generated lower performance. Participants in the high incentive treatments yielded significantly less points than those in the low incentive treatments. As standard theory and motivation crowding theory fail to give a convincing explanation for these results a model presenting reference dependent utility is applied.

An open question is why there was no similar outcome to Gneezy/Rustichini (2000a) not even with the IQ treatments. A possible explanation is the difference in wage composition namely the ratio between fix and variable components of the wage. Comparing Gneezy/Rustichini's (2000a) results to ours indicates that the composition of wages might have an essential influence on the impact of incentives and should be subject to further research.

Another interesting result is that the influence of specific task features seems to be (at least for our tasks) negligible since the outcome is qualitatively similar. A reason for that could possibly be that the differences between the tasks were too small to produce a measurable effect. The hypothesis of the IQ task generating more intrinsic motivation than the CN task cannot be confirmed. Neither the IQ task nor the CN task created a result, which could be interpreted as crowding out of intrinsic motivation. Nevertheless it is not understood that different tasks trigger off the same behavior. Still, the influence of task features cannot be resolved with these experiments.

The vital conclusion of the experiments described in this paper is not that incentives do not work, as it can be seen comparing the results of the no incentive and the low incentive treatments. Moreover the implication should be that incentives do work very well if the agent's income is below her reference income. Beyond this point however incentives may fail to provide more effort. Further evidence is necessary to confirm the hypothesis that participants, who were confronted with a higher variable pay, were quickly pleased with their wage. Nevertheless standard implications and the resulting practical advices must at least in some cases be doubted.

Still, a problematical question is what practical implications can actually be derived from the experiments, since they only admit suppositions on the participants' motives. Yet the individual wage expectations seem to be important information required to provide optimal incentives. Hence, this paper cannot give advice for the design of an optimal compensation scheme. For this purpose, further examinations of the emergence, development and measurement of reference points is necessary. The comparison with Gneezy/Rustichini (2000a) also indicates that the ratio between fixed payment and incentive parameter might as well play an essential role with motivating employees. Future research of this coherence might lead to interesting insights not only for the optimal height of incentives but also for the optimal composition of wages.

Appendix

Example for the task CN

This is the first page of the task CN translated from German

Please count the numbers of "Ones" and "Sevens"

Block 1

0	1	8	5	8	7	0	4	2	4	5	7	0	8	2	4	8	5	7	4	5	8	2	1	6	2	7	6	4	4
3	8	5	2	3	9	7	9	9	3	5	5	6	0	7	1	7	5	3	0	9	6	0	6	3	3	8	5	7	9
6	4	7	6	3	2	2	4	5	8	7	7	5	1	4	2	8	7	6	6	5	3	4	9	8	3	3	0	9	8
2	8	2	6	0	4	7	6	2	0	0	2	5	3	4	5	5	4	7	4	1	1	6	0	2	5	4	3	0	0
4	4	3	1	1	9	2	5	2	4	3	5	4	3	9	8	4	0	2	7	0	8	7	0	5	3	4	3	1	1
3	4	9	8	2	6	3	3	8	2	0	9	3	4	0	0	4	5	5	5	8	9	5	5	0	1	6	4	8	6
2	6	9	0	5	8	8	5	2	1	9	5	0	1	5	3	1	7	0	3	2	4	9	6	2	4	9	7	8	3
8	6	0	5	5	2	9	9	2	3	6	6	8	1	6	3	2	3	5	6	1	4	9	3	4	4	2	1	5	1
9	4	9	3	4	3	3	2	3	0	8	5	7	5	3	7	7	4	0	3	8	3	8	3	7	2	7	7	4	3
8	0	7	2	9	2	2	2	5	4	3	8	5	7	6	6	5	8	7	7	0	0	1	8	3	7	2	0	0	4
9	8	8	6	3	9	7	5	1	5	2	2	6	8	4	8	7	6	3	9	4	9	6	0	3	7	2	6	1	5
0	8	5	2	9	1	1	9	6	7	6	6	3	8	4	3	9	7	2	3	9	8	5	3	9	8	3	2	1	9

Number of Ones	
Number of Sevens	

Block 2

8	5	2	6	4	0	5	4	7	9	2	8	6	3	7	2	0	6	5	3	7	2	9	1	1	1	7	8	4	7
6	3	9	8	0	7	8	0	5	7	9	8	1	6	2	1	3	5	9	2	6	5	6	2	7	3	1	9	3	4
0	6	9	4	3	5	6	6	3	8	4	1	2	1	0	3	4	6	9	7	3	9	1	6	7	6	2	1	9	9
1	5	3	3	3	4	0	7	3	3	6	6	0	8	6	9	0	6	6	9	2	0	1	4	6	5	2	4	7	1
5	5	2	7	4	7	9	3	6	5	0	6	8	5	3	9	4	9	8	2	7	1	6	5	9	1	1	4	8	0
0	4	6	7	9	8	5	8	6	5	4	0	4	9	9	7	8	8	9	5	1	9	1	4	8	6	7	8	0	8
0	1	1	5	2	1	0	6	4	9	5	3	0	2	1	1	6	3	2	7	3	6	7	1	8	6	8	2	7	8
1	7	5	6	9	3	6	3	3	5	0	5	4	4	1	8	2	9	0	7	2	2	6	8	4	4	3	9	6	5
0	9	5	2	6	7	0	2	0	3	7	6	9	6	4	4	5	9	0	3	1	5	3	1	9	4	9	7	5	2
7	4	3	5	3	7	5	8	4	1	6	9	5	0	0	0	8	6	6	0	3	9	2	0	3	6	2	2	2	3
8	7	2	3	1	1	0	7	6	2	9	3	4	3	5	4	8	4	0	4	3	9	8	7	0	5	8	6	1	8
9	4	1	5	2	1	8	6	7	5	9	0	0	3	9	4	1	7	6	1	3	5	8	9	5	6	1	9	8	2

Number of Ones	
Number of Sevens	

Table 2

	Point Sorings					
	Bonn					
	Mean			**Median**		
Compensation	NI	LI	HI	NI	LI	HI
CN	23,40 (9,3)	26,07(9,0)	21,47 (7,4)	24	26	22
IQT	13,63 (9,3)	17,03 (8,0)	15,87 (9,2)	13	19,5	12,5
	Cologne					
	Mean			**Median**		
Compensation	NI	LI	HI	NI	LI	HI
CN	27,33 (10,9)	24,63 (7,9)	23,00 (7,3)	26	25	22,5
IQT	19,14 (6,5)	20,14 (10,2)	16,67 (10,4)	19,75	20	17,25

(Standard deviations in parentheses)

Table 3

Model of point scorings
Estimates from median regression
Dependent variable: Points

Explanatory variable	Regression coefficients (t-statistics)	
	Bonn	Cologne
Intercept	27,00***	26,00***
	(-14,74)	*(9,70)*
Fix	-4,00*	-1,00
	(-1,69)	*(-0,29)*
High	-5,00**	-5,50*
	(-2,19)	*(-1,67)*
IQ	-9,50***	-5,00*
	(-5,04)	*(-1,82)*
No. Of Observations	77	104
Pseudo R^2	0,1228	0,0655

For definition of the variables see Table 4.

Table 5

Models of points achieved
Estimates from OLS regression
Dependent variable:

Pooled Data

Explanatory variable	Regression coefficients (t-statistics)
	Model
Intercept	22,74
	(0,92)
Fix	-0,60
	(-0,37)
High	-3,57**
	(-2,26)
Cologne	1,79
	(1,33)
IQ	-5,79***
	(-4,35)
Sem	1,13
	(1,58)
Sem²	-0,11*
	(-1,94)
Age	0,29
	(0,15)
Age²	-0,01
	(-0,33)
No. Of Observations	181
Adjusted R²	0,16

For definition of the variables see Table 4.

Proof of Proposition

Due to the strict concavity of the objective function and the assumption on the cost function there must be a unique internal optimum. Suppose that $e^* < e^R$, then

$$e^* = c'^{-1}(\beta)$$

must be valid in the optimum. That is, the agent chooses an effort level, which is left of the kink. This occurs if and only if

$$\beta < c'\left(\frac{R-\alpha}{\beta}\right).$$

In case of $e^* > e^R$, the optimum is defined by

$$e^* = c'^{-1}(s\beta).$$

Consequently e^* must be located right of e^R which happens if

$$\beta > \frac{1}{s}c'\left(\frac{R-\alpha}{\beta}\right).$$

It follows that in all other cases the agent chooses $e^* = e^R$.

q.e.d.

References

Camerer, C., Babcock, L., Loewenstein, G. and Thaler, R., 1997. Labor Supply of New York City Cabdrivers: One Day at a Time. Quarterly Journal of Economics 112, 407-441.

DeCharms, R., 1968, Personal Causation: The internal affective determinants of behavior (New York Academic Press).

Deci, E. L., 1971. Effects of Externally Mediated Rewards on Intrinsic Motivation. Journal of Personality and Social Psychology 18, No.1, 105-115.

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

Deci, E. L., Koestner, R. & Ryan, R. M., 1999b. The undermining effect is reality after all-extrinsic rewards, task interest, and self-determination: reply to Eisenberger, Pierce, and Cameron (1999) Lepper, Henderlong, and Gingras (1999).
Psychological Bulletin 125, 692-700.

Dickinson, D. L., 1999. An Experimental Examination of Labor Supply and Work Intensities. Journal of Labor Economics 17, 638-670.

Easterlin, R. A., 2001. Income and Happiness: Towards a Unified Theory. The Economic Journal 111 (July), 465-484.

Eisenberger, R. & Cameron, J., 1996. Detrimental effects of reward. Reality or myth?.
American Psychologist 51 (11), 1153-1166.

Fehr, E. and Gaechter, S., 2001. Do Incentive Contracts Crowd Out Voluntary Cooperation?. Institute for Empirical Research in Economics, University of Zurich 2002, Working Paper No. 34.

Fehr, E. and Goette, L., 2002. Do Workers Work More if Wages are High? Evidence from a Randomized Field Experiment. University of Zurich 2002, Working Paper No. 125.

Frey, B. S., 1997, Markt und Motivation - Wie ökonomische Anreize die (Arbeits-) Moral verdrängen (Vahlen, München).

Gneezy, U. and Rustichini, A., 2000a. Pay Enough or Don't Pay at All. Quarterly Journal of Economics 115, 791-810.

Gneezy, U. and Rustichini, A.. 2000b. A Fine is a Price. Journal of Legal Studies 29, 1-18.

Irlenbusch, B. and Sliwka, D., 2004, Incentives, Decision Frames and Motivation Crowding Out-An Experimental Investigation. Mimeo, London School of Economics and Political Science and University of Cologne

Kahneman, D. and Tversky, A., 1979. Prospect Theory: An Analysis of Decision under Risk. Econometrica 47(2), 263-291.

Kunz, A. and Pfaff, D., 2002. Agency Theory, Performance Evaluation and the Hypothetical Construct of Intrinsic Motivation. Accounting, Economics and Society 27, 275-295.

Lazear, E. P., 2000. Performance Pay and Productivity. American Economic Review 90, 1346-1362.

MacCrimmon, K. R. and Larsson, S., 1979. Utility Theory: Axioms versus Paradoxes, In: M. Allais and O. Hagen (Eds.), Expected Utility Hypothesis and the Allais Paradox. Dordrecht/Boston, Reidel pp. 333-409.

Paarsch, H. J. and Shearer, B. S., 1999. The Response of Worker Effort to Piece Rates: Evidence from the British Columbia Tree-Planting Industry. Journal of Human Resources 34, 643-667.

Tversky, A. and Kahneman, D., 1991. Loss Aversion in Riskless Choice: A Reference Dependent Model. Quarterly Journal of Economics 106 (4), 1039-1061.