

# Reference Points and Bonus Payments – Field Evidence from a Multinational Firm

Axel Ockenfels\*                      Dirk Sliwka†  
University of Cologne              University of Cologne and IZA  
Peter Werner‡  
University of Cologne

Preliminary Version - Please do not circulate.  
February, 2009

## Abstract

We empirically investigate the effect of reference point violations on job satisfaction and performance of managerial employees in a large company in Germany and the United States.

We find evidence in line with loss aversion and inequality aversion. While negative deviations from a reference point induced by the system have significant detrimental effects on satisfaction, the impact of positive deviations is nil. Furthermore, violations of managers' reference points affect performance negatively.

**Key Words:** Reference points, loss aversion, incentives, bonus payments, job satisfaction

**JEL Classification:** J33, M52

---

\*University of Cologne, Albertus-Magnus-Platz, D-50923 Köln, Germany. tel.:+49 221 470-5761, fax: +49 221 470-5068, e-mail: ockenfels@uni-koeln.de.

†University of Cologne, Herbert-Lewin-Str. 2, D-50931 Köln, Germany, tel: +49 221 470-5888, fax: +49 221 470-5078, e-mail: dirk.sliwka@uni-koeln.de.

‡University of Cologne, Albertus-Magnus-Platz, D-50923 Köln, Germany, tel.:+49 221 470-4354, fax: +49 221 470-5068, e-mail: peter.werner@uni-koeln.de

# 1 Introduction

Relative positions of employees affect wage structures in firms. In an often-cited survey study of more than 300 US companies, Bewley (1999) concludes that “within a company, pay inequity offends (indeed, sometimes outrages) employees and destroys trust”. There is evidence that employees not only orientate on absolute wages for their evaluations of payments but also on relative standings.

The importance of relative comparisons is well established in happiness research: Subjective well-being and reported life satisfaction are highly related to relative wealth positions (for economic applications of these phenomena see Frank & Sunstein (2000) and Clark et al. (2007)). There is also a growing empirical and experimental literature suggesting an impact of relative comparisons on satisfaction and (real or experimental) performance of subjects in labor market environments (see Pfeffer & Langton (1993), Brown et al. (2008) and Torgler et al. (2008) for empirical studies and Greenberg (1988) and Clark et al. (2006) for experiments). Generally, these studies document detrimental effects of wage/payout dispersion or status inequalities that are usually driven by negative reactions of persons in inferior positions.

In economic models, the impact of relative positions have been incorporated in various ways. Akerlof & Yellen (1990) proposed a negative relationship between a worker’s effort and her wage relative to a visualized ‘fair’ wage. If realized wages are lower than the wages perceived as fair, effort exertion of employees is undermined. Moreover, the existence of positional concerns among workers, i.e. the dependency of utility on one’s rank in a reference population, may lead to wage compression in firms (see Frank (1984*a*), Frank (1984*b*)).

The importance of relative payoffs in a labor market environment can be motivated also by theories of inequality aversion (Bolton & Ockenfels (2000) and Fehr & Schmidt (1999)). These models assume that utility or motivation does not depend solely on absolute payoffs of a person, but also on her payoff standing relative to a social reference point or to other persons.

Finally, loss aversion plays a role in this context. The concept of loss aversion postulates that utility is evaluated relative to a reference point. Losses, i.e. negative deviations from a given reference point, decrease utility stronger than same-sized positive deviations increase utility (see Kahne-

man & Tversky (1979), Tversky & Kahneman (1991) and Kahneman et al. (1991)).

The identification of the impact of loss aversion in labor market environments is, however, difficult. For given incentive schemes, there are multiple reference points that could possibly influence employees. For this reason, related theoretical and empirical evidence is scarce. Köszegi & Rabin (2006) proposed a model of reference-dependent preferences in which reference points are formed on the basis of rational expectations and apply this concept among other things to labor supply decisions. In an empirical study, Mas (2006) showed that police performance in the US was sensitive to pay rises compared to reference points set by final offer arbitrations. If arbitrators in a compensation dispute decided in a way unfavorable to the demand of the police force, the number of crimes cleared decreased afterwards.

There is a large literature on the optimal design of incentive schemes (see Prendergast (1999) for a survey of incentive theories). Empirical studies show a generally positive impact of variable pay on performance (see Lazear (2000) and Bandiera et al. (2007)). However, variable payment schemes may not provide optimal incentives if agents have social preferences. Sliwka (2007) introduced a model in which fixed rather than variable wages chosen by the principal send a positive signal about her trust in a cooperative social norm and may induce higher effort exertion. For a literature overview about the impact of social preferences on different incentive schemes see Englmaier (2005).

To our knowledge, there is no study investigating the impact of reference points for incentive schemes. In business practise, substantial importance is attached to the design of variable pay systems. However, we are not aware of an empirical test about the importance of reference points for the success of these systems.

We analyze the impact of reference points both on job satisfaction and performance with the example of a typical bonus system for executives of a world-wide operating company. The connection of survey, compensation and performance data on the individual level enables us to assess incentive and satisfaction effects of given system features and their implementation in practice. Moreover, the implemented system has the advantage that it provides a clear reference point against which managers can evaluate their bonus payments.

We find that - in line with loss aversion and inequality aversion - negative deviations from the reference point have significant detrimental effects on reported job satisfaction while the impact of positive markups is nil. Furthermore, lack of transparency mitigates the importance of reference points, and spot bonuses have a stronger effect on satisfaction than regular bonus payments even though their economic relevance for employees is much smaller. With respect to performance, we find that violations of managers' reference points affect their supervisor performance negatively. The negative effect is robust against alternative specifications of reference points, evaluation behavior and team composition. Section 2 describes the survey and the compensation data; section 3 presents the results for reported job satisfaction and performance. In section 4, we provide a brief discussion of our results and conclude.

## **2 The Data**

### **2.1 Survey, Compensation and Performance Data**

We study if social preferences and reference points affect employees' perceptions of payment systems. These questions are addressed using survey and compensation data from managers of a multinational company (>100,000 employees) operating in diverse business fields. In cooperation with the company, we conducted a survey among managers located in Germany (autumn 2007) and the United States (summer 2008). The large majority of managers of the company are subject to the same system of variable compensation (the main exception being sales-related positions). However, the handling of the system differs in small but important details that are decisive for the formation of reference points and incentives induced by the system.

As the survey was part of a larger study, managerial employees were asked some 60 questions about job satisfaction and motivation, workplace characteristics, and preferences for incentive schemes. Together with the survey, we collected personal data about demographic characteristics, department affiliations, performance evaluation and compensation over the years 2004-2006 (Germany) and 2004-2007 (US). The technical environment allowed us to connect the background data with survey answers in a way that guaranteed anonymity of the participants. Altogether 4,997 executives

took part in the survey (3,122 in Germany and 1,875 in the US).<sup>1</sup>

Contrary to measures of satisfaction and motivation directly derived from the survey, objective performance measures are not available due to diversity of tasks between and within company divisions. Moreover, in some cases the characteristics of the task itself makes an objective evaluation difficult (for example for managers in the research department). For this reason, we rely on subjective performance evaluation measures for our analysis of incentives.

## 2.2 Characteristics of the Bonus System

At the end of every business year, individual bonus budgets are determined by linking company performance, hierarchy level and fixed salary of a manager. Each supervisor receives the sum of individual bonus budgets to distribute among the managers in her department. Then, she allocates personal payout percentages (markups on or reduction below 100% bonus budget).

This decision depends on the individual performance evaluation of a person. The system defines payout percentage intervals for each of five possible performance ratings (including ‘Excellent’, ‘Above Average’, ‘Average’, ‘Below Average’ and ‘Inadequate’). For example, a person rated ‘Average’ (the majority of managers) should be assigned between 80% - 110% of her individual bonus budget. If a supervisor evaluates all her managers with ‘Average’ ratings, she is able to allocate every person the actual budget of 100%. However, the interrelation between performance ratings and payout percentages makes bonus assignments a zero-sum game. If a supervisor differentiates in ratings, externalities are created by the budget: If one manager receives a bonus payment  $>100\%$  of her budget, another manager must necessarily receive a bonus  $<100\%$ . Moreover, the variety of hierarchical levels expands the budgeting pressure for the supervisors.

Bonus payments of German and US managers are roughly comparable during the period of analysis. Average bonus payments of German managers account for some some 21,800 USD in 2006 while US managers received

---

<sup>1</sup>In the remainder of the paper, we use only data sets for which it was possible to connect survey answers with demographic and compensation data. For a substantial share of managers this connection was not feasible, for example due to gaps in the data or due to the fact that a person has entered a manager level only recently.

Table 1: Distribution of Performance Ratings

	D 2006	US 2007	Recommended
‘Excellent‘	0.6	2.0	$\leq 5\%$
‘Above Average‘	22.4	33.9	$\leq 25\%$
‘Average‘	74.1	62.7	$\sim 60\%$
‘Below Average‘	2.8	1.4	$\leq 10\%^*$
‘Inadequate‘	0.1	n/a	

\*Recommended proportion refers to the combined share of ‘Below Average‘ and ‘Inadequate‘ managers.

some 22,500 USD in 2007 (the years relevant for the survey). Furthermore, average payouts remain largely constant over the years in the sample.<sup>2</sup>

Given the connection between ratings and bonuses, an important aspect is the practice of performance evaluation. Table 1 lists the proportions of performance ratings in the year of our survey. The actual shares of performance ratings are contrasted to the recommended distribution provided by the company.<sup>3</sup>

We find ratings biases often discussed in the literature of subjective performance evaluation. Leniency biases (the tendency to rate employees too positive) and centrality biases (compression of ratings) are well established empirical phenomena (see Prendergast (1999) for a survey of related studies).<sup>4</sup> In Germany, differentiation of ratings is less pronounced than desired by the company, as the large majority of managers receives the medium rating ‘Average‘. There is only a very small share of extraordinary ratings or grades below average (in sum less than 5%). Supervisors in the US on average do not differentiate strongly either. Yet their evaluation behavior is different: In general, American supervisors assign the highest performance ratings more often whereas there are less cases of lower-than-average ratings. Furthermore, there is the tendency towards a leniency bias, as the share of ‘Above Average‘ ratings is substantially higher than recommended.<sup>5</sup>

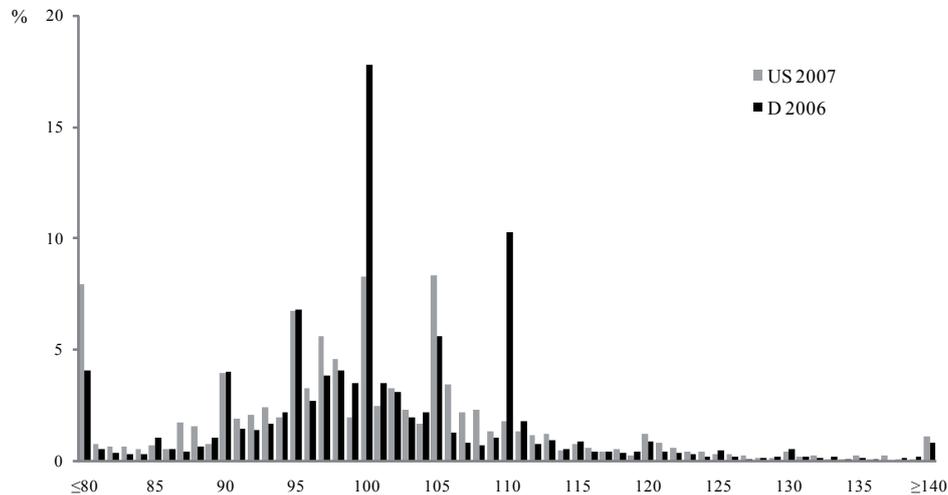
<sup>2</sup>It is important, however, that there is a substantial spread of bonus payments between managers on different hierarchy levels.

<sup>3</sup>Please note that the recommended distribution has the character of a decision support for supervisors and is not a forced distribution.

<sup>4</sup>There is also evidence suggesting that supervisor evaluation biases also influence subordinates‘ subsequent performance (see the recent study by Bol (2008)).

<sup>5</sup>The leniency bias could be partly explained by different interpretations of the ratings in the countries. In interviews with responsible supervisors in the US it was stated that

Figure 1: Percent Shares of Average Bonus Payout Percentages (Regular Bonus)



The combination of company success, individual performance evaluation and budget allocation finally leads to bonus payouts. Figure 1 shows the distribution of realized bonus payouts in Germany and the US. Although the bonus system enables supervisors to differentiate strongly between employees (theoretically it is possible to assign between 0% and 160% of the bonus budget), realized payouts are strongly compressed. The large majority of payout percentages lies around 100% of the budget. While the distributions are in general similar in Germany and the US, an important difference is that US supervisors are more likely to assign more distinct payout percentages (less than 80% and more than 120% of the budget). Generally, the distribution reflects evaluation patterns in Germany and the US (remember that each performance rating is associated with an interval of payout percentages).

Several percentage values serve obviously as focal points for bonus allocation, the most prominent being 100% of the bonus budget.<sup>6</sup> In Germany, nearly 20% of all managerial employees are paid out exactly their budgets. This focal point is far less decisive in the US, however.

---

the rating ‘Below Average’ was associated with more severe implications for the further career of the respective managers than in Germany. Thus, US supervisor might be more likely to refrain from assigning this rating.

<sup>6</sup>A further strong focal point in Germany is 110%, the lower threshold for managers rated ‘Above Average’.

A fundamental aspect for our analysis is that the mode of communication of the bonus payments differs across countries in a way that enables us to assess the economic impact of reference points. In Germany, managers receive information highlighting their personal payout percentages together with the notice about their yearly bonus payment. Contrary, when US managers receive the notice about their bonuses, only absolute amounts are communicated.<sup>7</sup> This difference will have large impacts on the managers' perception and incentive effects induced by the system, as will be shown in the remainder of the paper.

Finally, in addition to regular bonus payments, supervisors can allocate individual spot bonuses to employees, among other things for special achievements or exceptional performance in particular projects. However, there exist no formal allocation rules, and the practice of spot bonus allocations differs strongly between supervisors. In most cases, less than 20% of the managers per supervisor receive spot bonuses. Compared to regular bonus payments, the economic relevance of these payments is low: The sum of spot bonuses accounts for some 3% of total bonus payments; on average spot bonuses of managers who are assigned a payment account for 1.5 - 2.5 % of their yearly incomes.

### 3 Results

In the discussion of our results, we will first concentrate on the determinants of individual job satisfaction. The assessment of the effects of reference point violation will be in focus. For this purpose we use cross-section regression models, in which we investigate the impact of individual compensation and performance parameters.

Next, a crucial question is whether reference points also affect incentives for performance. Using personal data from 2004-2006 (Germany) and from 2004-2007 (US), supervisor performance ratings are our proxies for performance. Due to the panel structure of our data set, we are able to track a supervisor's evaluation behavior in a given year and the evolution of her performance in the subsequent year over 2(3) years in our sample.

---

<sup>7</sup>Please note that US managers are in general able to infer individual payout percentages. However, the connection between actual bonus payment and payout percentage is not explicitly communicated and emphasized so strongly as in Germany.

### 3.1 Reference Points and Job Satisfaction

Our proxy variable for job satisfaction is the answer to the question: „I am very satisfied with my job.“ Subjects could respond on a scale from 1 (totally disagree) to 7 (fully agree), so that increasing values of the score reflect higher satisfaction. The distributions of answers in the US and Germany are similar as table 2 shows.

Table 2: Proportions of answers to the question "I am very satisfied with my job." in %

	D 2006	US 2007
1 - totally disagree	1.0	1.1
2	3.0	3.9
3	6.0	6.6
4	10.2	13.9
5	27.9	27.5
6	40.0	34.4
7 - fully agree	11.9	12.7

To analyze the impact of compensation instruments on individual job satisfaction, we estimate regression models with job satisfaction as the dependent variable. In line with Freeman (1978), we use a unit normal transformation of the satisfaction score by subtracting the sample mean from a person's score and dividing the result by the standard deviation of the sample.

In the first step, we investigate the impact of compensation instruments on reported satisfaction in the year of the survey. We calculate models of the form

$$JobSatisfaction_{it} = \alpha + \beta \cdot X_{it} + \varepsilon_{it}$$

where  $X_{it}$  is the matrix of compensation, performance evaluation variables as well as controls for demographics, organizational backgrounds and supervisor evaluation behavior. To account for the discrete outcomes of the dependent variable, we calculate ordered probit models on the cross-section of employee data in the year 2006 (Germany) and 2007 (US), respectively.

Table 3 lists alternative specifications for the satisfaction models. The first important thing to notice is that - controlling for the specific background of a manager - there is no correlation between salary variables and

reported satisfaction (model 1 for D and US). In neither country we find a significant impact of the absolute salary or the salary increase in the preceding year. Furthermore, - despite their substantial economic relevance - absolute bonus payouts have no significant influence either. The respective coefficients even have negative signs. This latter finding supports the argument that monetary short-term incentives on their own are not fully suited to induce motivation and increase job satisfaction among employees.

By contrast, the performance rating significantly and robustly influence reported satisfaction of a person. In specification 1, we include dummy variables for the performance ratings with ‘Average’ as the reference group. For both countries, regression coefficients have the expected signs: Better-than-average ratings are positively associated with satisfaction while the opposite is true for lower-than average ratings.<sup>8</sup> This finding is robust to alternative specifications and indicates the decisive impact of performance evaluation for reported satisfaction. A possible reason suggested by the data is that evaluation ratings not only influence short-term bonus payouts but also long term career opportunities.<sup>9</sup>

Finally, although absolute amounts are much lower than regular bonus payments, spot bonuses are significantly associated with a higher satisfaction score in case of Germany. This is in line with the results of Engellandt & Riphahn (2004) indicating a positive relation between surprise bonuses and proxies for employee effort. In the US, the coefficient of spot bonuses is also positive but insignificant.

In the next step, we focus on the role of payout percentages and the threshold of 100% budget. As external factors such as company success and hierarchy levels strongly affect absolute bonus amounts, we expect the perceived influence of the individual to be low. Therefore, in specification 2 we include payout percentages instead of absolute amounts. Indeed, we find payout percentages to be positively and significantly correlated with reported satisfaction. The more a supervisor assigns a manager from her individual budget, the higher is the satisfaction score.

However, following the literature about the effect of reference point vio-

---

<sup>8</sup>The missing significance of ‘Excellent’ and ‘Inadequate’ ratings in Germany is probably related to the small numbers of managers in the sample who were assigned these ratings.

<sup>9</sup>Generally we find a positive correlations between personal performance ratings and the probability of promotions to higher hierarchy levels in several regression specifications.

Table 3: Determinants of Job Satisfaction (Dependent Variable: Adjusted Satisfaction Level)

Country No. Model	D 1 Ordered Probit	D 2 Ordered Probit	D 3 Ordered Probit	US 1 Ordered Probit	US 2 Ordered Probit	US 3 Ordered Probit
Fixed Salary (000s)	0.005 [0.006]	0.003 [0.005]	0.004 [0.005]	-0.003 [0.004]	-0.003 [0.003]	-0.003 [0.003]
Salary Increase (000s)	0.007 [0.005]	0.005 [0.005]	0.006 [0.005]	-0.024 [0.016]	-0.025 [0.016]	-0.025 [0.016]
Spot Bonus (000s)	0.041* [0.022]	0.042** [0.022]	0.043** [0.022]	0.023 [0.026]	0.031 [0.026]	0.031 [0.026]
Performance Rating ‘Excellent’	0.405 [0.281]			0.667** [0.263]		
Performance Rating ‘Above Average’	0.235*** [0.067]			0.248*** [0.083]		
Performance Rating ‘Below Average’	-0.920*** [0.168]			-1.462*** [0.364]		
Performance Rating ‘Inadequate’	-0.616 [0.763]					
Absolute Bonus Payout (000s)	-0.006 [0.010]			-0.004 [0.006]		
Bonus Payout Percentage		0.012*** [0.002]			0.007** [0.003]	
Positive Deviation from Bonus Budget (=100%)			0.003 [0.004]			0.006 [0.005]
Negative Deviation from Bonus Budget (=100%)			-0.023*** [0.004]			-0.008* [0.004]
Observations	2093	2098	2098	917	918	918
log-likelihood	-3175	-3192	-3186	-1403	-1416	-1416

Standard errors are given in brackets. \*, \*\* and \*\*\* denote significance on the 10%, 5% and 1%-level, respectively. Demographic Control Variables include a manager’s age and gender, total years of affiliation to the company and at a given hierarchy level as well as dummies for company divisions, hierarchy levels and promotions in the last year.

Control Variables of supervisor evaluation behavior include coefficients of variation of performance ratings and bonus payout percentages.

lations, we suspect that the relation of payout percentages and satisfaction is non-linear. Here, we exploit the fact that the system provides the clear reference point of 100% bonus budget for each individual. In model 3, we split payout percentages into those above and those below 100% bonus budget and measure deviations of the actual payout from the threshold. This model is defined as

$$\begin{aligned} \text{JobSatisfaction}_{it} = & \\ & \alpha + \beta \cdot X_{it} + \gamma \cdot (STI\%_{it} - 100\%) \cdot \mathbf{I}_{\{STI\%_{it} \geq 100\%\}} + \\ & \delta \cdot (100 - STI\%_{it}) \cdot \mathbf{I}_{\{STI\%_{it} < 100\%\}} + \varepsilon_{it} \end{aligned}$$

where  $\mathbf{I}_{\{STI\%_{it} \geq 100\%\}}$  and  $\mathbf{I}_{\{STI\%_{it} < 100\%\}}$  are dummy variables taking the value of 1 if the STI percentage is larger than/equal to 100% or below 100%, respectively. If reference points influence satisfaction in our sample, we expect  $\gamma$  and  $\delta$  to be different from each other. In particular, loss aversion would imply that  $|\gamma| < |\delta|$ .

Our models confirm patterns of reference point sensitivity in the case of Germany. Markups on 100% do not increase self-reported satisfaction significantly. The coefficient is positive but very small. Contrary, decreases below 100% have a significant negative effect on satisfaction levels: the farther away a person's payout percentage is from the 100% threshold, the lower is her estimated job satisfaction.

On first sight, the model for US managers provides similar evidence. Here, we find also a (weakly) significant effect of negative deviations from 100%, and no effect if the manager receives more than her budget.

However, given the strong correlation between performance evaluation and payout percentages, we have to rule out the possibility that the presumed reference point effect is not in fact driven by ratings. We control for this possibility by rerunning model 3 but include only managers whose payout percentages are located around the reference point (those with an 'Average' rating, see table 5 in the appendix). In the case of Germany, the reference point effect remains robust and highly significant. For the US, however, the significance of the negative deviation variable diminishes. Hence, reference point sensitivity is only found among German managers. Given the different levels of transparency with respect to payout percentages, this result is plausible. As in Germany the personal bonus budget is an open and prominent reference point, its violation is also reflected in

individual satisfaction. Contrary, US managers do not orientate on their personal budgets, and negative deviations from the personal budget do not lower their satisfaction accordingly.

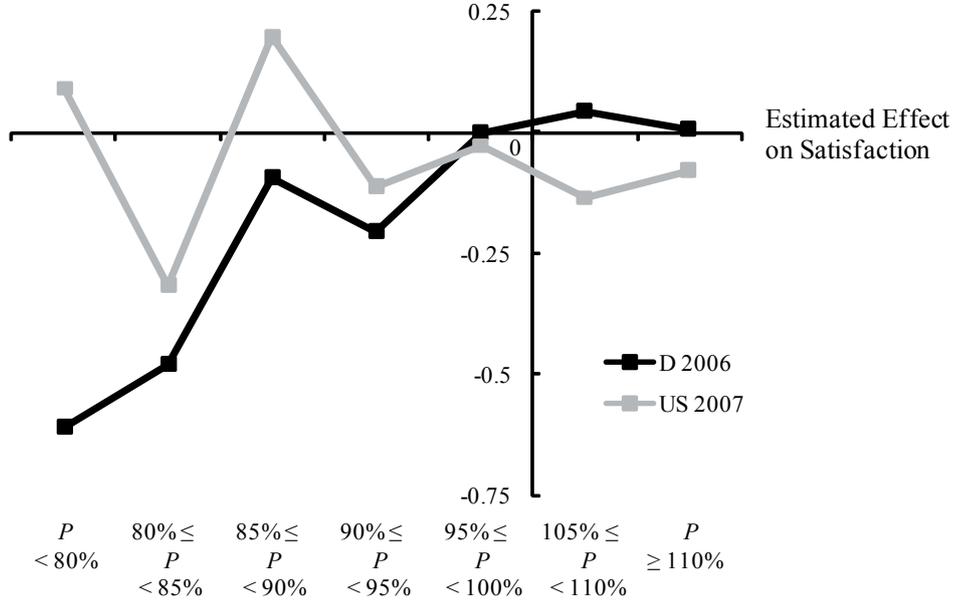
To confirm and to quantify this result, we calculate an additional OLS model with adjusted satisfaction as the dependent variable (table 6 in the appendix). Instead of payout percentage values we include dummies for percentage intervals. The reference group consists of managers who received payout percentages  $\geq 100\%$  and  $< 105\%$ . For German managers, dummies below 100% have negative signs and are significant in most cases, indicating a lower satisfaction score compared to managers around the threshold (the exception is the dummy for payouts larger than 95% where the coefficient is roughly zero). Moreover, the size of the coefficients tend to become larger the farther away the payout is from the reference point. Dummies on the positive side of the reference point exhibit positive and insignificant signs. Corresponding to the low importance of the reference point in the US, there is no clear tendency found in the model, and all payout dummies are insignificant. The estimated effects from the model are depicted in figure 2. In the case of Germany, the coefficients of the interval dummies exhibit roughly the kinked shape around the reference point implied by loss aversion. For the US sample, there is no clear relation found between percentage interval and job satisfaction.

All in all, our results have shown that if the system handling creates a reference point (here by the mode of communication), it influences satisfaction in the asymmetric way loss aversion suggests: Positive deviations from 100% are associated with small and insignificant increases of satisfaction. Contrary, negative deviations from 100% lead to satisfaction decreases that tend to become stronger, the further a manager moves away from 100%. Given this effect, the next question is whether reference point violation also translate into economic behavior, in our case the performance of the managers.

### **3.2 Reference Points and Performance**

So far, we cannot assess the impact of evaluation practise and bonus assignments on actual employee performance. Monetary incentive systems are implemented to induce effort exertion, and it is per se not clear if employee satisfaction and effort are correlated. To address the issue of incentive ef-

Figure 2: Estimated Effects of Payout Percentages on Reported Satisfaction



The figure shows the estimated effect of payout percentage intervals on job satisfaction, controlling for compensation, performance rating, and demographic background of the managers (see model 5 in the appendix). Managers with payout percentages equal or larger than 100% and smaller than 105% are the reference group.

facts, we rely on proxy measures, because direct controls are not available for the present company. We assume that a supervisor's rating in a given year - irrespective of her function - is positively correlated with the effort provided by her subordinates. Thus, performance ratings of the supervisors are our measures for the performance of the respective teams. We further assume that evaluation behavior of a supervisor in a given year affects motivation (and subsequently performance) of the team members in the following year. This corresponds to the following model with a supervisor's performance in year  $t + 1$  as the dependent variable

$$Rating_{it+1} = \alpha + \beta \cdot DevRef_{it} + \gamma \cdot X_{it} + \varepsilon_{it}$$

where  $X_{it}$  is a vector of individual background variables of a supervisor in the previous year  $t$  as well as a year dummy. As in the case of the satisfaction variable, we use a unit normal transformation of the perfor-

mance rating. Besides organizational affiliations and hierarchy levels, the background variables include compensation data (fixed salary and salary increase) and evaluation behavior in year  $t$ . For the latter, we use the shares of employees rated ‘Excellent’, ‘Above Average’ and ‘Below Average’.<sup>10</sup>

It is important to note that evaluation behavior and actual performance of the team are interdependent. First, the practice of bonus and rating assignments provides incentives for performance in the subsequent period. Our hypothesis is that, given the negative impact of reference point violation on satisfaction, supervisors who push a larger share of their subordinates below 100% exhibit relatively lower performance ratings than other supervisors in the subsequent year. To identify the reference point effect, we include several variables  $DevRef_{it}$  in some specifications indicating how strong reference points of subordinates have been violated.

However, a possible negative effect found in regressions could possibly be driven by the composition of teams. If a supervisor had only low-performing subordinates in her team, she would be forced to assign low payout percentages and - given the correlation between her performance and team performance - her rating should also be low. To distinguish between team composition and incentive effects, we compare the results of ordered probit models and linear models with fixed effects on the level of the supervisors. While the first class of models enables us to identify mainly general sample effects (team effects), the latter models give an insight how behavior of a particular supervisor influences incentives to perform well in the subsequent period. Table 4 lists models for Germany and the United States.

Model 1 serves as the benchmark for our analysis. In the ordered probit specifications, the share of assigned ratings have the expected signs: The share of ‘Excellent’ and ‘Above Average’ managers in the team is positively and significantly associated with the supervisor rating in the subsequent year. Contrary, the share of ‘Below Average’ managers affects performance negatively (significantly so in case of the US). However, an important difference between countries is found if the effects of evaluation behavior are analyzed on the supervisor level. While the assignment of the best possible grade is positively related to supervisor performance in both countries, the sign of ‘Below Average’ ratings differs. In Germany, supervisor ratings are

---

<sup>10</sup>We include no dummy for the rating ‘Inadequate’, because it is hardly ever assigned in the data set.

Table 4: Incentive Effects of Deviations from Reference Points (Dependent Variable: Adjusted Performance Rating of Supervisor in Subsequent Year)

Country No. Model	D		D		D		US		US	
	Ordered Probit	Fixed Effects								
Share of Employees rated 'Average' and awarded below 100%	0.01	0.065*	-0.491*	-0.825**	0.007**	-0.033	0.330	0.211	0.007**	-0.031
	[0.007]	[0.036]	[0.277]	[0.403]	[0.003]	[0.022]	[0.236]	[0.278]	[0.003]	[0.022]
Fixed Salary	0.017*	0.065*	0.011	0.062*	0.031**	0.026*	0.030**	0.025*	0.031**	0.014
	[0.009]	[0.012]	[0.009]	[0.012]	[0.012]	[0.014]	[0.012]	[0.014]	[0.012]	[0.014]
Salary Increase	3.423*	6.199**	3.498*	6.745**	2.918***	2.086*	3.131***	2.240**	2.918***	2.240**
	[1.939]	[2.795]	[1.939]	[2.768]	[0.816]	[1.104]	[0.831]	[1.125]	[0.816]	[1.104]
Share of Employees rated 'Above Average'	1.374***	-1.238**	1.252***	-1.352***	0.647**	-0.064	0.903***	0.095	0.647**	-0.064
	[0.305]	[0.483]	[0.313]	[0.479]	[0.290]	[0.373]	[0.343]	[0.429]	[0.290]	[0.373]
Share of Employees rated 'Below Average'	-0.191	3.167**	-0.517	2.727*	-10.056***	-4.153**	-9.873***	-3.950**	-10.056***	-4.153**
	[1.079]	[1.467]	[1.095]	[1.461]	[2.492]	[1.956]	[2.503]	[1.978]	[2.492]	[1.956]
Constant	3.469*	3.469*	3.990*	3.990*	5.318	5.318	4.874	4.874	5.318	5.318
	[2.077]	[2.077]	[2.063]	[2.063]	[3.256]	[3.256]	[3.314]	[3.314]	[3.256]	[3.256]
Observations	367	367	367	367	297	297	297	297	297	297
R-squared (within)	.	0.17	.	0.2	.	0.18	.	0.19	.	0.19
log-likelihood	-444.6	.	-443	.	-486.8	.	-485.9	.	-486.8	.
Number of Supervisors	249	249	249	249	176	176	176	176	176	176

Standard errors are given in brackets. \*, \*\* and \*\*\* denote significance on the 10%, 5% and 1%-level, respectively. Demographic Control Variables include the age of the supervisor, total years of her affiliation to the company, and correction for part-time employment.

higher, the higher the share of these managers is, indicating that differentiation in ratings generally has a positive effect on incentives and ultimately on performance. For the US, the opposite is true: the higher the share of managers with a bad rating, the worse the performance rating of the respective supervisor. This result possibly reflects the above-mentioned ‘stigma’ effect of this rating. Assigning a ‘Below Average’ may have a negative incentive effect on the whole team.

Furthermore, the share of ‘Above Average’ managers is negatively associated with performance in Germany. The reason for this observation could lie in the budgetary pressure associated with this rating. Supervisors with a high share of positively rated managers are forced by the system to fund their payouts by lowering payout percentages of the remaining managers in their team. Hence, this result can be interpreted as a first indication for a detrimental effect of violation of reference points. In the US, however, the budgetary effect is weaker, as will be argued in the following.

Finally, compensation variables have significant positive effects in some model specifications, the most robust being the positive impact of salary increases on the level of a supervisor in the US. Here, the data suggests the functioning of financial incentives, as stronger salary increases are correlated with higher performance in the subsequent year.

Given the effects of background variables, we analyze the implications of reference points. Our hypothesis that violations are detrimental is supported in model 2, where we include the share of managers per supervisor rated ‘Average’ and receiving less than 100% bonus payout. Across countries we find results in line with the regressions on reported satisfaction. Controlling for background, compensation data and evaluation behavior of the supervisor, reference point violations negatively affect supervisor performance in Germany. In specification 2, the control variable is negatively associated with a supervisor’s performance. Thus, the more employees are pushed below the reference point with their bonus payments in a given year, the lower is estimated performance in the next year. This effect is robust both in ordered probit models and on the level of particular supervisors. There are two interpretations of this result: First, for a given team composition, good supervisors realize that one should take reference points into account for the evaluation. Second, a given supervisor demotivates her managers when violating reference points. Furthermore, as in model 1, the share

of ‘Above Average’ managers is still negatively related to performance of a given supervisor. Alternative specifications of the model 2 with an interaction variable of the share of ‘Above Average’ managers and the share of ‘average’ managers below 100% suggests that the negative effect of the former is at least partly driven by budgetary pressure.

Contrary to the German models, we do not find a significant detrimental effect of reference point violations in the US data. In model 2 (and also in the other specifications), the share of ‘Average’ managers below 100% even exhibits a positive sign. As managers do not orientate strongly on the reference point of 100% budget, performance is not negatively affected. Here, supervisors can achieve a stronger variation in payouts avoiding negative impacts of reference point violations. Moreover, the absence of the reference point even enables supervisors to allocate more high ratings to their subordinates, because budgetary pressure is lower (in the US, some 11% more of the managers receive ‘Above Average’ than in Germany).<sup>11</sup>

In the following, we describe additional robustness checks to validate the previous statements. The effects of reference point violations in Germany (and the lack of an effect in the US) remain robust in alternative specifications. Table 7 in the appendix lists models using further measures for team composition and reference point violation.

First, we insert the mean and the coefficient of variation of assigned ratings instead of relative frequencies in the regressions (models 3). The reference point effects remain equally significant in Germany, and overall conclusions from evaluation behavior are similar to specification 1.

Second, we include mean payout percentages per supervisor for ‘Average’ managers as an alternative measure for reference point violation (see models 4). Also in this specification, the effect remains equally robust. The respective coefficient is positive and significant for Germany (both in the ordered probit and the fixed effect model), indicating a stronger reduction in payouts of ‘Average’ managers decreases incentives accordingly.

Finally, our main result could be driven by financial constraints of the supervisor rather than behavioral effects of reference point violation. Therefore, we control for absolute bonus budgets per supervisor (models 5). The

---

<sup>11</sup>Please note that managers rated ‘Average’ in the US receive about 3% less bonus payout. Among supervisors who differentiate strongly in performance ratings, this difference even increases to more than 7% on average.

less budget per capita a supervisor is able to distribute, the more severe would be budget effects, and the more managers would have to be reduced below the threshold for a given performance rating distribution. However, the inclusion of bonus budgets does not have a significant impacts for both countries, and the effect of reference point violation remains significant for Germany.

Summarizing, our data provides evidence that reference point violation does not only affect employee perception but also has an impact on performance. As in the case of satisfaction, our results highlight the impacts of differences in system handling: Only where communication fosters orientation on reference points, detrimental effects of its violations are found. An otherwise identical spread of payout percentages is not found to influence performance if the reference point is not salient.

## 4 Discussion and Conclusion

We have conducted an empirical study in which we connect demographic, compensation and performance data with survey answers to derive insights about the relevance of reference points for variable payment schemes. Comparing managerial employees in the United States and Germany, we find effects consistent with loss aversion in the latter country. Negative payout deviations from the reference point induced by the system significantly lower reported job satisfaction, whereas positive deviations have little influence. The absolute size of bonus payoffs does not have a significant impact on satisfaction. Moreover, a violation of reference points influences performance, as supervisor ratings drop with higher shares of managers being assigned a bonus below their budget. In the US, system handling differs in a way that prevents the formation of reference points, as personal payout percentages are not explicitly communicated. Subsequently, the described detrimental effects of reference point violations are not observed.

It is important to note that the exhibited patterns are not only in line with loss aversion but also with theories of inequity aversion. In the present bonus system, a reference point based on rational expectations and a social reference point coincide: First, 100% of the bonus budget is a natural payout expectation for ‘Average’ performance. At the same time, bonus assignments also transmit information about the reference group. Receiving

less than the budget, the manager knows that she has been rated lower and received proportionally less than other managers. Thus, any bonus assignment below the 100% threshold is equivalent to an inferior relative position among managers rated by the same supervisor.

In general, our study provides evidence that social preferences are reflected in employee satisfaction and also in performance. Because the effects are partly related to the handling of the system, we can infer some practical considerations for the design of variable payment schemes.

First, our data suggests that transparency of bonus payouts is a double-edged sword. Economic theory does not provide clear predictions about the optimal level of transparency within a given incentive system. Transparency may help to foster traceability and also the acceptance of bonus payouts. Yet, transparency can be detrimental if employees' expectations are violated and relative comparisons are triggered. Therefore, lower transparency might mitigate relative comparisons. This is also in line with experimental findings suggesting that a lack of transparency may dampen social comparison effects (see Gehrig et al. (2007)). At the same time, however, low transparency provides supervisors more degrees of freedom in the distribution of payouts. In our sample, US supervisors are able to differentiate more in bonus payments, because they do not have to orientate so strongly on reference points of their subordinates. Still, this leads to the undesired effects of higher arbitrariness of bonus assignments, and, as budgetary pressure is lower, performance ratings that are biased upwards.

The second and related point is how to differentiate in performance ratings and payouts. In our sample, stronger differentiation in ratings on the whole tends to have positive incentive effects. However, if reference points play a role, further differentiation within a given rating is detrimental. Due to behavioral patterns in line with loss aversion and inequity aversion, satisfaction gains or higher incentives on the positive side of the reference point are overcompensated by the losses in satisfaction and incentives on the negative side.

However, comparable studies with larger company data sets are necessary to get a broader overview on the interaction of social preferences and incentive system design. As empirical and experimental evidence clearly shows that employees do not respond solely on material rewards, the goal from a practical perspective is to design variable payment schemes that pro-

vide an adequate balance between material incentives and non-monetary preferences of employees.

## 5 Appendix

### References

- Akerlof, G. A. & Yellen, J. L. (1990), ‘The Fair Wage-Effort Hypothesis and Unemployment’, *Quarterly Journal of Economics* **105**, 255–283.
- Bandiera, O., Barankay, I. & Rasul, I. (2007), ‘Incentives for managers and inequality among workers: Evidence from a firm-level experiment’, *The Quarterly Journal of Economics* **122**(2), 729–773.
- Bewley, T. F. (1999), *Why Wages Don't Fall During a Recession*, Harvard University Press, Cambridge.
- Bol, J. C. (2008), ‘The determinants and performance effects of supervisor bias’, Working Paper.
- Bolton, G. E. & Ockenfels, A. (2000), ‘ERC - A Theory of Equity, Reciprocity and Competition’, *American Economic Review* **90**, 166–193.
- Brown, G. D. A., Gardner, J., Oswald, A. J. & Qian, J. (2008), ‘Does wage rank affect employees’ well-being?’, *Industrial Relations* **47**(3), 355 – 389.
- Clark, A. E., Frijters, P. & Shields, M. (2007), ‘Relative income, happiness and utility: An explanation for the easterlin paradox and other puzzles’, IZA Discussion Papers No. 2840.
- Clark, A. E., Masclet, D. & Villeval, M.-C. (2006), ‘Effort and comparison income: Survey and experimental evidence’, PSE Working Paper No. 2006-03.
- Engellandt, A. & Riphahn, R. T. (2004), ‘Incentive effects of bonus payments: Evidence from an international company’, IZA Discussion Papers No.1229.
- Englmaier, F. (2005), A Survey on Moral Hazard, Contracts, and Social Preferences, *in* B. Agarwal & A. Vercelli, eds, ‘Psychology, Rationality and Economic Behaviour: Challenging Standard Assumptions’, Palgrave Macmillan, pp. 125–139.

- Fehr, E. & Schmidt, K. (1999), ‘A Theory of Fairness, Competition, and Cooperation’, *Quarterly Journal of Economics* **114**, 817–868.
- Frank, R. H. (1984a), ‘Are workers paid their marginal products?’, *American Economic Review* **74**, 549–571.
- Frank, R. H. (1984b), ‘Interdependent preferences and the competitive wage structure’, *The Rand Journal of Economics* **15**, 510–520.
- Frank, R. H. & Sunstein, C. (2000), ‘Cost benefit analysis and relative position’, AEI-Brookings Joint Center For Regulatory Studies, Research Paper No. 00-5.
- Freeman, R. B. (1978), ‘Job satisfaction as an economic variable’, *American Economic Review* **68**(2), 135–41.
- Gehrig, T., Guth, W., Levati, V., Levinsky, R., Ockenfels, A., Uske, T. & Weiland, T. (2007), ‘Buying a pig in a poke: An experimental study of unconditional veto power’, *Journal of Economic Psychology* **28** (6), 692–703.
- Greenberg, J. (1988), ‘Equity and workplace status: A field experiment’, *Journal of Applied Psychology* **73**(4), 606–613.
- Kahneman, D., Knetsch, J. L. & Thaler, R. H. (1991), ‘The endowment effect, loss aversion, and status quo bias: Anomalies’, *Journal of Economic Perspectives* **5**(1), 193–206.
- Kahneman, D. & Tversky, A. (1979), ‘Prospect theory: An analysis of decision under risk’, *Econometrica* **47**(2), 263–91.
- Köszegi, B. & Rabin, M. (2006), ‘A model of reference-dependent preferences’, *Quarterly Journal of Economics* **121**(4)(4), 1133 – 1165.
- Lazear, E. (2000), ‘Performance pay and productivity’, *American Economic Review* **90**(5), 1346–1361.
- Mas, A. (2006), ‘Pay, reference points, and police performance’, *Quarterly Journal of Economics* **121**(3)(3), 783–821.
- Pfeffer, J. & Langton, N. (1993), ‘The effect of wage dispersion on satisfaction, productivity, and working collaboratively: Evidence from college and university faculty’, *Administrative Science Quarterly* **38**, 382–407.

- Prendergast, C. (1999), 'The provision of incentives in firms', *Journal of Economic Literature* **37(1)**, 7 – 63.
- Sliwka, D. (2007), 'Trust as a signal of a social norm and the hidden costs of incentive schemes', *American Economic Review* **97(3)**, 999–1012.
- Torgler, B., Schaffner, M., Schmidt, S. L. & Frey, B. S. (2008), 'Do employees care about their relative position? behavioural evidence focusing on performance', CREMA Working Paper Series 2008-12.
- Tversky, A. & Kahneman, D. (1991), 'Loss aversion in riskless choice: A reference-dependent model', *The Quarterly Journal of Economics* **106(4)**, 1039–61.

Table 5: Determinants of Job Satisfaction among Employees rated ‘Average’  
(Dependent Variable: Adjusted Satisfaction Level)

Country No. Model	D 4 Ordered Probit	US 4 Ordered Probit
Fixed Salary (000s)	0.003 [0.006]	-0.004 [0.004]
Salary Increase (000s)	0.005 [0.006]	0.002 [0.025]
Spot Bonus (000s)	0.049* [0.028]	0.041 [0.047]
Positive Deviation from Bonus Budget (=100%)	-0.001 [0.012]	-0.009 [0.010]
Negative Deviation from Bonus Budget (=100%)	-0.024*** [0.007]	-0.006 [0.006]
Observations	1544	574
log-likelihood	-2364	-897

Standard errors are given in brackets. \*, \*\* and \*\*\* denote significance on the 10%, 5% and 1%-level, respectively. Demographic Control Variables include a manager’s age and gender, total years of affiliation to the company and at a given hierarchy level as well as dummies for company divisions, hierarchy levels and promotions in the last year.

Control Variables of supervisor evaluation behavior include coefficients of variation of performance ratings and bonus payout percentages.

Table 6: Determinants of Job Satisfaction (Dependent Variable: Adjusted Satisfaction Level)

Country Model	D OLS	US OLS
Fixed Salary (000s)	0.003 [0.005]	-0.004 [0.003]
Salary Increase (000s)	0.003 [0.005]	-0.022 [0.015]
Spot Bonus (000s)	0.030 [0.020]	0.024 [0.024]
Performance Rating ‘Excellent’	0.237 [0.275]	0.565** [0.268]
Performance Rating ‘Above Average’	0.140 [0.113]	0.284** [0.120]
Performance Rating ‘Below Average’	-0.538* [0.309]	-1.591*** [0.370]
Performance Rating ‘Inadequate’	-0.218 [0.804]	
Payout < 80%	-0.608* [0.363]	0.093 [0.184]
80% ≤ Payout < 85%	-0.478*** [0.141]	-0.314 [0.262]
85% ≤ Payout < 90%	-0.092 [0.126]	0.2 [0.215]
90% ≤ Payout < 95%	-0.204** [0.084]	-0.11 [0.149]
95% ≤ Payout < 100%	0.000 [0.064]	-0.025 [0.119]
105% ≤ Payout < 110%	0.045 [0.087]	-0.133 [0.134]
Payout ≥ 110%	0.007 [0.122]	-0.076 [0.148]
Constant	-0.281 [0.287]	-0.185 [0.313]
Observations	2093	917
R-squared	0.09	0.07

Standard errors are given in brackets. \*, \*\* and \*\*\* denote significance on the 10%, 5% and 1%-level, respectively. Demographic Control Variables include a manager’s age and gender, total years of affiliation to the company and at a given hierarchy level as well as dummies for company divisions, hierarchy levels and promotions in the last year. Control Variables of supervisor evaluation behavior include coefficients of variation of performance ratings and bonus payout percentages.

