Bonus payments and reference point violations

Axel Ockenfels
University of Cologne

Dirk Sliwka
University of Cologne and IZA

Peter Werner
University of Cologne

First version: 17 December 2009
This version: 10 March 2011

Abstract
We investigate how bonus payments affect satisfaction and performance of managers in a large, multinational company. We find that falling behind a naturally occurring reference point for bonus comparisons reduces satisfaction and subsequent performance. The effects are mitigated if information about one’s relative standing towards the reference point is withheld. A model and a laboratory experiment provide complementary insights and additional robustness checks.

Key Words: Reference points, bonus payments, job satisfaction, job performance, transparency

JEL Classification: D03, M52

* Corresponding author: Axel Ockenfels, University of Cologne, Albertus-Magnus-Platz, D-50923 Köln, Germany (e-mail: ockenfels at uni-koeln.de). We thank the editor, three anonymous referees, David Jaeger and seminar audiences at Aarhus, Bonn, Cologne, Dortmund, Edinburgh, Granada, Heidelberg, Karlsruhe, London, Magdeburg, Mannheim, Munich and Toulouse for very helpful comments that greatly improved our paper. We also thank the board and many managers of the studied company for sharing their data and views with us. Ockenfels thanks the German Science Foundation for financial support.
This paper studies the effects of bonus payments on satisfaction and performance. By merging compensation and evaluation data from personnel records of a large company with survey responses, we find that falling behind a natural reference point for a fair bonus (a “reference point violation”) significantly decreases both, satisfaction and subsequent performance of the affected managers. The data also suggest that, because of the detrimental effects of reference point violations, the distribution of bonuses gets compressed at the reference point. However, the effects are substantially mitigated if a manager’s relative standing towards the reference point is made less transparent.

Potentially harmful effects of reference point violations in work settings have been noticed before. Regarding satisfaction, Truman F. Bewley (1999, p.43) concludes from his surveys covering more than 300 US firms that “within a company, pay inequity offends (indeed, sometimes outrages) employees and destroys trust.” Regarding performance, George Akerlof and Janet Yellen (1990) introduce the “fair wage-effort hypothesis”, postulating that workers proportionately withdraw effort as their actual wage falls short of the fair wage. Yet, so far, there has been only little evidence on the impact of reference-dependent preferences and pay transparency on satisfaction, performance and payment patterns based on naturally occurring field data. One reason is that, without experimental control, it is often difficult to identify the ‘right’ frame of reference and its transparency to workers in labor relationships.\(^1\)

We make use of a unique data set to study the causes and effects of reference point formation in the context of a bonus plan for managers in a multinational company. One important feature of our data is that the company’s bonus system produces a clear reference point, largely consistent with behavioral economics approaches as discussed below. Also, our data allow us to match demographic characteristics, salaries, and bonus

\(^1\) Intriguing studies that investigated the impact of reference points in a natural field environment are Alexandre Mas (2006), who shows that police performance is sensitive to pay rises compared to reference points set by final offer arbitrations, and Illoong Kwon and Eva Meyersson Milgrom (2009), who analyze exogenous changes in workers’ relative wages during M&As and find that status in the workplace affect turnover decisions. David Card et al. (2010) and Yan Chen et al. (2010) study the effects of transparency about one’s relative standing with the help of controlled field studies. Others employ laboratory experiments to study reference point and transparency effects. We will later get back to this literature.
payments from personnel records with survey responses of managers in Germany (where the headquarters are located) as well as the US. Finally, the data include reference point violations under two different degrees of transparency of the reference standard.

The company’s bonus system stipulates that each year each manager is assigned a bonus target (also called the “bonus budget”). The individual target depends on company performance, divisional performance and the manager’s salary grade, among other variables. The bonus payment is set by the supervisor as a percentage of the individual manager’s target (the so called “bonus percentage”). A manager’s bonus percentage is restricted by his performance rating, which is determined and communicated several weeks before the individual target is calculated and his bonus percentage is assigned. The performance rating is chosen from a five point rating scale (‘excellent’, ‘above average’, ‘fully meets expectations’, ‘below average’, and ‘inadequate’). By far most managers’ rating is ‘fully meets expectations’, and the bonus percentage of these managers must be chosen between 80% and 110%. Better rated managers must be assigned 110% or more, and worse rated managers must be assigned less than 80%. Also, the bonuses paid to all managers in a department are limited such that the sum of bonus payments cannot exceed the sum of all individual bonus budgets.

Within this context, the 100% bonus percentage is a natural reference point for managers. In fact, it is in line with the reference points proposed by two widely-used types of reference-dependent preferences. The first type comprises social comparison models such as Ernst Fehr and Klaus Schmidt (1999) and Gary E. Bolton and Axel Ockenfels (2000), which imply that people dislike falling behind others. In our context, given the fixed budget for total bonus payments within a department, the department’s supervisor always has the option to choose an equal distribution in the sense of paying all managers in her department the respective budget (that is, a bonus percentage of 100%). Yet if she

---

2 The company has more than 100,000 employees worldwide. The largest subsidiaries are located in Germany and the US. Our study covers the incentive scheme for all managers in both countries. The term “manager” refers to executives and to all other employees not covered by a collective wage agreement. In this paper we refer to managers as those who are assigned bonuses, and to “supervisors” as those who assign bonuses. Managers in upper hierarchical levels can also be supervisors (who receive and assign bonuses), something we will exploit in our analyses in Section III.
decides to give one manager more, others must get less. Therefore, falling behind the 100% bonus percentage implies that others get a share of one’s own budget, leading to socially unfavorable inequality measured in bonus percentage terms. Hence, the 100% bonus percentage is a natural social comparison standard for bonus payments.

Based on the seminal work of Daniel Kahneman and Amos Tversky (1979), the second type of relevant reference-dependent preferences typically deal with one-person decision problems, in which outcomes are evaluated relative to the status quo. Botond Köszegi and Matthew Rabin (2006) further develop the essential intuition of Kahneman and Tversky’s work, postulating that people dislike falling behind one’s expectation held recently about outcomes. Applied to our context, a manager who was rated ‘fully meets expectations’ is likely to interpret this as hitting the target, and so is likely to expect a bonus of (at least) 100%. Thus, the 100% bonus percentage is both a reasonable social reference point and a reasonable expectation for ‘fully meets expectations’ managers.

One common prediction of both types of reference-dependent preferences is that falling below the respective reference standard has a larger effect on satisfaction than a same-sized gain above the reference point. In social contexts, Bolton (1991) and Fehr and Schmidt (1999), among others, discuss evidence and present theory suggesting that disadvantageous inequality hurts more than advantageous inequality, which might even often be irrelevant altogether. Similarly, the models by Kahneman and Tversky (1979) and Köszegi and Rabin (2006) imply that losses are more powerful than gains. So, we hypothesize that assigning a manager less than 100% bonus percentage will strongly negatively affect a manager’s satisfaction, while assigning more yields – if at all – a small effect.

Regarding the effect of reference point violations on the managers’ performance, there are competing hypotheses. On the one hand, there is literature suggesting that reference point violations can decrease performance. Most prominently, Akerlof and Yellen (1990) motivate their fair wage-effort hypothesis (that reference point violations reduce effort) with research in psychology and sociology and as it “accords with common sense”. More recent behavioral economics research suggests that if reference point violations are
perceived as an unkind act, managers may reciprocally respond with unkindness, and decide to reduce effort levels in the future (Rabin, 1993, Martin Dufwenberg and Georg Kirchsteiger, 2004, Armin Falk and Urs Fischbacher, 2006). On the other hand, however, one might think that (the possibility of) falling below the reference point creates additional incentives to improve one’s performance and to avoid falling below one’s standard. More specifically, the company’s bonus system resembles a tournament, because a total department’s bonus is limited, and the bonus percentages rank workers according to performance. Tournament theory suggests that compensating workers according to rank may increase effort (Edward P. Lazear and Sherwin Rosen, 1981).

Summing up, theory does not yield unambiguous predictions regarding the effect of reference point violations on performance. Thus, while we hypothesize that falling below the reference point strongly diminishes satisfaction, we refrain from formulating a hypothesis about how reference point violations affect performance. However, we expect that any satisfaction and performance effects due to reference-dependent preferences that we observe in Germany are mitigated in the US. The reason is that while the bonus schemes are basically identical, bonus percentages are explicitly communicated to managers in Germany but not to managers in the US. This is because labor regulations in Germany require a higher degree of transparency in compensation systems. As bonus percentages cannot easily be inferred from absolute bonus payments and other data available to US managers (see Section I), we hypothesize that the lack of transparency in the US system diminishes the role of the 100% reference point.

---

3 Indeed, many laboratory and some field studies, starting with Fehr, Kirchsteiger and Arno Riedl (1993), suggest that reciprocity can play an important role in labor relationships. See David Cooper and John Kagel (forthcoming) for a recent survey on other-regarding preferences, and Simon Gächter and Fehr (2002), who survey laboratory labor market research. For more recent evidence see Fehr, Lorenz Götte and Christian Zehnder (2009) and the references therein. Also, there is laboratory evidence that subjects who fall below a reference point, be it the status quo (Kahneman and Tversky, 1979) or a social reference point (Bolton and Ockenfels, 2010), are willing to take more risks, which may additionally affect performance. Other related experimental studies find that wage comparisons affect behavior in experimental labor relations (see our discussion in Section IV).

4 Dominique Demougin and Claude Fluet (2003) and Christian Grund and Dirk Sliwka (2005) show in theoretical models that the incentive effect of tournaments may become stronger when individuals are inequity averse.
Overall, our data strongly support the key predictions of reference-dependent preferences and of Akerlof and Yellen’s (1990) fair wage-effort hypothesis. We find that violations of the managers’ 100% reference point significantly negatively affect both satisfaction and future performance in Germany, where reference point violations are transparent. We also find that bonuses are significantly more compressed at the reference point in Germany than in the US. At the same time, we find no reference-dependent satisfaction and performance effects for the company’s managers in the US, who operate under a less transparent but otherwise identical bonus system as their colleagues in Germany.

Section I describes our data and shows that the distribution of bonus percentages in Germany is much more compressed at 100% in Germany than in the US. Section II analyzes the determinants of job satisfaction and establishes a strong and robust relevance of the 100% reference point for managers’ satisfaction in Germany. It also sketches a model that straightforwardly adds reference-dependent preferences to a standard model of subjective performance evaluation. Our model organizes the observed interaction between transparency of the reference standard on the one hand and satisfaction and the compression of bonuses on the other hand. Section III investigates the impact of reference point violations on performance. In Section IV, we provide a further robustness check of our main results in a controlled laboratory environment that captures the relevant features of the bonus system. Moreover, the experiment complements the indication in the field data that the performance effect is partly due to negative reciprocity. Section V discusses the findings and concludes.
I. A first look at the managers’ compensation

Supported by the board of the company, we conducted a survey among the managerial staff in Germany (autumn 2007) and the US (summer 2008). As the survey was part of a larger study, managers were asked some 60 questions about workplace characteristics and other work-related aspects, including job satisfaction. We independently collected data about performance evaluations, compensation, demographic characteristics, and department affiliations of the managers over the years 2004-2006 (Germany) and 2004-2007 (US) from the personnel records of the firm. The technical environment allowed us to connect this background data with the survey answers of the individual managers in a way that guaranteed anonymity of the participants of the survey. 5 Altogether 4,997 managers took part in the survey (3,122 in Germany and 1,875 in the US), which corresponds to a participation rate of 59% and 41%, respectively.6

The company has several main units working in different industries (each of which consists of a large number of divisions), and each unit is present both in Germany and the US. When the present bonus system was implemented, the company’s goal was to align procedures everywhere in order to maximize cross-country comparability, mobility and flexibility. As a result, Germany and the US have common organizational structures, and the bonus systems in Germany and the US operate under basically identical formal rules. There are also no systematic differences in informal rules, for example with respect to goal-setting, or in the communication content, timing and procedures – with one

5 The procedure guaranteed that no party involved in data generation and processing could combine the information about contact details, background data and survey answers of the participants. First, we collected the administrative data from the personnel records of the company and encoded the data set. Then the company received the encoded data set, added email addresses of all managers and transferred the data to an independent consulting firm who administered the survey. This firm then invited the managers by email to take part in the survey, which was conducted in the intranet of the firm with an anonymous code, and deleted the email addresses afterwards. Finally, we received the encoded data set including survey answers without information that allowed identification of individuals. Managers were informed on the first page of the survey that their answers would be matched anonymously with their individual compensation and background data and had to agree to this procedure to participate in the survey.

6 Comparing participating and non-participating managers, we do not find significant differences in demographic or compensation variables, so that we have no reason to suppose that the representativeness of the sample is limited. The lower response rate in the US can be attributed to the fact that the US survey was run at a time when a larger fraction of the employees was on holiday.
important exception: Due to tighter transparency regulations, managers in Germany, when receiving their yearly bonus letters, are explicitly told their bonus percentage together with their bonus payments. In the US, on the other hand, managers are only told their bonus payment.  

Salaries and absolute bonus payments of German and US managers are roughly comparable during the period of analysis. In 2006 (2007), the average fixed salary of a manager in Germany (in the US) was 107,700 US-Dollars (113,200 US-Dollars). The bonus payments are substantial, as they account for some 20% of fixed salaries on average (22,200 US-Dollars in Germany and 22,700 US-Dollars in the US).

Figure 1. Distribution of bonus percentages in Germany and the US

---

7 Individual bonus budgets are computed based on salary grade and several performance indicators measuring the financial success of the company as a whole as well as of the relevant division, subject to exceptions and discretion. Moreover, bonus budgets may vary substantially (the yearly growth rate of the divisional performance measure in the relevant time window varies between – 4.4% and 26.4%). These complexities make it generally very difficult for a US manager to learn his individual budget. In fact, the open survey answers of US managers show that some managers wish to get better informed. Typical comments include: “The letter that comes with the award should clearly spell out your rating”, and “At a minimum, the employee should be told what % was applied”.

8 While US managers receive somewhat larger base salaries, the share of managers with bonuses less than 10,000 US-Dollars is higher in the US than in Germany (18.9% versus 5.7% of the managers in the sample). This is because the percentage share of variable pay is 2-7% lower for US managers on the lowest hierarchy level.
However, there are first indications of the importance of the 100% reference point when we look at how bonuses in Germany and the US are distributed. Figure 1 shows the bonus percentage distributions in Germany for 2006 and in the US for 2007, the respective years of our survey (the distributions are very similar in the other years).

Compared to the US data, bonus percentages in Germany are much more compressed towards the 100% reference point. Most strikingly, in Germany, there is a significant peak at 100%, with 17.6% of the managers receiving exactly their target bonus, while in the US only 8.3% of the managers receive a 100% bonus percentage ($p < 0.001$, two-sided $\chi^2$-test). The next largest peak in Germany is at 110%, which is the feasible bonus closest to 100% for the ‘above average’ managers. In the US, on the other hand, 110% exhibits no specific prominence – only 4.3% of all ‘above average’ managers in the US receive exactly 110%, which is about one eighth of the corresponding number in Germany (36.1%; $p < 0.001$). Also, the share of managers in Germany with bonus percentages below 80% (above 110%) is with 2.2% (11.6%) compared to 6.6% (13.4%) in the US (weakly) significantly smaller at $p < 0.001$ ($p = 0.054$). Looking at the whole distribution, the standard deviation of bonus percentages in Germany is with 12.2 also significantly smaller than the standard deviation of 15.1 observed in the US ($p < 0.001$, two-sided Levene test for differences in variances).

We summarize our findings in:

**Observation 1.** The distribution of bonus percentages is more compressed towards the 100% reference point in Germany, where bonus percentages are made transparent, than in the US, where managers do not learn their bonus percentages. In particular, there is a significant fraction of managers receiving 100% in Germany, while the corresponding frequency is less than half as high in the US.

As we will explain in more detail later, we attribute the differences in the distributions of bonus percentages to the different degree of transparency regarding the reference point in Germany and the US. Because pushing managers below 100% reduces satisfaction and performance if and only if the reference point violations are transparent (see Sections II
and III), there is more reason for compression towards the reference point for ‘fully meets expectation’ managers in Germany than in the US (as predicted by our model described in Section II). A similar argument holds for ‘above average’ managers. The more managers are pushed above 100%, the larger the budgetary pressure to violate other managers’ reference points, which is harmful when the violations are transparent. In fact, we observe that the performance rating ‘above average’ is assigned significantly less often in Germany (see Table A1 in the Appendix for the distribution of the performance ratings) and, if it is assigned, the bonuses of these managers are often 110%, which minimizes the pressure to violate the reference point of other managers.

Our evidence and interpretation is in line with other studies that associate compressions in payments with reference points and social concerns. In particular, it is consistent with Robert H. Frank’s (1984) seminal work showing that pay compression, relative to worker productivity, can be explained as the result of workers caring for their relative position (see also Frank 1985), as well as with Akerlof and Yellen’s (1990) fair wage-effort hypothesis.

II. How does the bonus matter for job satisfaction?

In this section, we investigate whether there is a kink in the managers’ preferences at the 100% reference point as postulated by the models of reference-dependent preferences outlined in our introduction. While we cannot measure preferences directly, our study design allows us to investigate how deviations from the reference point affect perceived job satisfaction. A commonly used proxy for job satisfaction is the degree of approval to the statement: “I am very satisfied with my job.” In our survey, managers could respond to this question on a scale from 1 (totally disagree) to 7 (fully agree), so that increasing values of the score reflect higher satisfaction levels. The distributions of

---

9 In line with this notion, US supervisors tend to deviate more from a manager’s individual budget: In the years relevant for our survey, the average share of ‘fully meets expectations’-managers per supervisor who receive bonuses lower than their budgets is 81.9% in the US compared to 61.0% in Germany.

responses in the US and Germany are rather similar, with a mean of 5.29 (standard deviation 1.25) in Germany and 5.24 (1.25) in the US (see Table A2 in Appendix A). We estimate ordered probit regressions with job satisfaction as the dependent variable for the years 2006 (Germany) and 2007 (US), in which the respective survey studies took place. We include controls for age intervals, gender, firm tenure, promotion in the previous business year and the company unit. In some specifications we also control for the effects of performance ratings. Table 1 reports the regression results.

In our first specification, we focus on the impact of absolute bonus payments and bonus percentages. As a manager’s bonus budget depends on his position in the company, salary, and on the performance of both, the company and his division, bonus payments and bonus percentages are not collinear, so that the effect of the bonus payment can be disentangled from that of the bonus percentage. In Model 1, the coefficient of the absolute bonuses is insignificant, while the coefficient of bonus percentages is highly significant and positive for both Germany (Model D1) and the US (Model US1). Hence, despite their substantial economic relevance, absolute bonus payments do not have a robust influence on job satisfaction – while bonus percentages do.

---

11 The results are similar when we use OLS regressions on a unit normal transformation (see for instance Freeman, 1978) of the satisfaction scores.
12 The correlation coefficient is $r = 0.204$ in the German and $r = 0.153$ in the US sample.
13 Absolute bonus payments and bonus percentages have both a positive and mostly significant impact when only one variable is included in the regression. Also, the regressions do not consistently show a significant positive relation between the salary in the year of the survey and self-reported satisfaction. This evidence is in line with a body of research suggesting that an increase in already high income levels does not substantially affect satisfaction levels (see, for instance, Clark, 1999, and the references cited therein).
Table 1. Determinants of job satisfaction
(Ordered Probit models with the ‘job satisfaction’ score as dependent variable)

<table>
<thead>
<tr>
<th></th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>D5</th>
<th>US1</th>
<th>US2</th>
<th>US3</th>
<th>US4</th>
<th>US5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salary (000s)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.002</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.005</td>
</tr>
<tr>
<td></td>
<td>[0.003]</td>
<td>[0.003]</td>
<td>[0.003]</td>
<td>[0.004]</td>
<td>[0.003]</td>
<td>[0.003]</td>
<td>[0.003]</td>
<td>[0.003]</td>
<td>[0.003]</td>
<td>[0.004]</td>
</tr>
<tr>
<td>Bonus Payment (000s)</td>
<td>0.001</td>
<td>0.000</td>
<td>-0.002</td>
<td>-0.002</td>
<td>-0.006</td>
<td>0.001</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.011**</td>
</tr>
<tr>
<td></td>
<td>[0.009]</td>
<td>[0.009]</td>
<td>[0.009]</td>
<td>[0.010]</td>
<td>[0.004]</td>
<td>[0.004]</td>
<td>[0.004]</td>
<td>[0.004]</td>
<td>[0.004]</td>
<td>[0.006]</td>
</tr>
<tr>
<td>Bonus Percentage</td>
<td>0.016***</td>
<td>0.008**</td>
<td>0.012***</td>
<td>-0.002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.003]</td>
<td>[0.004]</td>
<td>[0.004]</td>
<td>[0.006]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dummy ‘Above meets’</td>
<td>0.143*</td>
<td>0.270***</td>
<td>0.233**</td>
<td>0.232**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.082]</td>
<td>[0.090]</td>
<td>[0.111]</td>
<td>[0.111]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dummy ‘Below meets’</td>
<td>-0.791***</td>
<td>-0.420*</td>
<td>-1.123***</td>
<td>-1.150***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.189]</td>
<td>[0.217]</td>
<td>[0.355]</td>
<td>[0.436]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive Deviation from Bonus</td>
<td>0.005</td>
<td>-0.005</td>
<td>0.004</td>
<td>0.005</td>
<td>-0.001</td>
<td>-0.015</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.004]</td>
<td>[0.005]</td>
<td>[0.012]</td>
<td>[0.006]</td>
<td>[0.008]</td>
<td>[0.026]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative Deviation from Bonus</td>
<td>-0.032***</td>
<td>-0.025***</td>
<td>-0.027***</td>
<td>-0.019***</td>
<td>0.003</td>
<td>0.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.005]</td>
<td>[0.006]</td>
<td>[0.007]</td>
<td>[0.006]</td>
<td>[0.010]</td>
<td>[0.012]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Budget (=100%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample</td>
<td>All</td>
<td>All</td>
<td>All</td>
<td>All</td>
<td>All</td>
<td>‘fully meets’</td>
<td>All</td>
<td>All</td>
<td>All</td>
<td>All</td>
</tr>
<tr>
<td>Observations</td>
<td>2045</td>
<td>2045</td>
<td>2045</td>
<td>2045</td>
<td>1507</td>
<td>956</td>
<td>956</td>
<td>956</td>
<td>956</td>
<td>599</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-3073</td>
<td>-3064</td>
<td>-3064</td>
<td>-3058</td>
<td>-2276</td>
<td>-1455</td>
<td>-1450</td>
<td>-1454</td>
<td>-1450</td>
<td>-927</td>
</tr>
</tbody>
</table>

Standard errors are given in brackets. *, ** and *** denote significance on the 10%, 5% and 1%-level, respectively. Control variables include age interval dummies, gender, total years of affiliation to the company, a dummy for a promotion in the previous year, and dummies for company units.
When we include rating dummies (Model 2), the described effect remains significant for Germany, yet the US coefficient of the bonus percentage becomes statistically as well as economically insignificant. As managers in the US do not learn their bonus percentages, they cannot infer their position within the percentage interval corresponding to a given rating. Hence, controlling for the rating, one’s relative position towards the reference point does not matter for satisfaction. The effect of percentages observed in model US1 is therefore entirely driven by performance grades. In Germany however, one’s relative standing within a given performance rating is known to managers and, hence, significantly correlates with job satisfaction.14

As we have pointed out in our introduction, one key characteristic of reference-dependent preferences is that people dislike falling below the reference point more than they like exceeding the reference point by the same amount. Therefore, we hypothesize that reference point violations – getting a bonus percentage below 100% – significantly decrease job satisfaction, whereas increasing the bonus beyond 100% does not or only marginally affect satisfaction. Model 3 tests this hypothesis. In this specification we allow for the possibility that the effect of the bonus percentage has different slopes below and above the reference point:

\[
JobSatisfaction_i = \alpha + \beta \cdot X_i + \gamma \cdot (z_i - 100) \cdot I_{z_i > 100\%} + \delta \cdot (100 - z_i) \cdot I_{z_i < 100\%} + \varepsilon_i,
\]

where \(z_i\) is the bonus percentage of individual \(i\) and \(I_{z_i > 100\%}\) (\(I_{z_i < 100\%}\)) is a dummy variable taking the value 1 if the bonus percentage is above 100% (below 100%). Hence, \(\gamma\) captures the effect of a positive deviation and \(\delta\) of a negative deviation from the 100% bonus percentage. If 100% constitutes a reference point relative to which employees evaluate their bonus, we expect that \(\delta\) is significantly negative, and \(\gamma\) is positive but small.

14 There is a strong connection between performance ratings and job satisfaction in Germany as well as in the US: Managers with a rating better than ‘fully meets expectations’ are significantly happier, and those with a rating worse than that grade are significantly less happy.
Consistent with the hypothesis we find that falling behind the 100% bonus significantly decreases satisfaction levels, while increases above 100% do not increase self-reported satisfaction in Germany (see Model D3). The asymmetric effect of being below and above the reference is also apparent in the US (Model US3) – even though US managers do not know their bonus percentages. The reason is that bonus percentages are correlated with performance ratings, so US managers can use their ratings as a proxy of their relative standing. E.g., a manager rated ‘above average’ (‘below average’) knows that he must have a bonus percentage above (below) 100%. However, when controlling for the effects of performance ratings (Model 4), the asymmetric effect of deviations from 100% budget vanishes in the US sample while remaining robust in Germany.\(^{15}\)

We also estimate Model 3 with a sample including only managers who receive the ‘fully meets expectations’ rating. The result again remains stable in the German sample but disappears in the US (see D5 and US5).

Following our discussion in the introduction, the statistical analyses above assumed that the reference point (i.e., the kink in the estimated function) is at 100%. In order to further justify our choice and to also allow for different shapes, we ran simple OLS regressions on a unit normal transformation of the satisfaction score for Germany and the US, including dummies for percentage intervals instead of bonus percentages and controls for performance ratings as well as supervisor fixed effects (see Table A4 in Appendix A). The reference group consists of managers who receive exactly their budgets \((z_i = 100\%)\).

In the German sample, both dummies for intervals below 100% have statistically significant negative signs, indicating a lower satisfaction score compared to managers at the 100% threshold. Both dummies for intervals above 100% are positive, statistically insignificant and small in size. In the US, too, both dummies below (above) 100% have

\(^{15}\) The results are not due to potential ceiling effects of the job satisfaction variable: If we exclude all managers with satisfaction scores of 7, so that truncation cannot be an issue in the remaining sample, our conclusions are the same. See Table A3 in the Appendix. Moreover, the reported effects of reference point violations are equally valid if we use ‘satisfaction with the bonus payment’ (measured on the same 1-to-7 scale as ‘job satisfaction’) as the dependent variable.
negative (positive) signs, but here all interval dummies are insignificant. Figure 2 illustrates the results. For Germany, estimated interval dummies exhibit the expected kinked shape at the reference point. No such pattern is observed in the US data.

Figure 2. Relation between bonus percentage and job satisfaction relative to managers with a 100% bonus percentage

The figure shows the estimated values for bonus percentage interval dummies on a unit normal transformation of the job satisfaction score, controlling for compensation, performance rating, and demographic background of the managers and including fixed effects per supervisor (see Table A4 in the Appendix). Significant interval dummies are marked with an asterisk (*).

16 The 100% reference point is further justified when running OLS regressions with the same set of variables as in Models D3, D4 and D5, but varying the position of the hypothesized kink between 90% and 110% in steps of 1%. It turns out that the goodness-of-fit measured by the R-squared (within) value is maximized with a kink at 100 or 101%. Following the Kahneman-Tversky tradition, one might also argue that the bonus in the previous year is another plausible candidate for a reference point. However, we did not find evidence for this hypothesis. We believe that part of the reason is that the individual bonus budget strongly depends on company and divisional financial performance which are both very volatile (see footnote 7). Hence, last year’s bonuses cannot easily serve as a reference point for expectations regarding subsequent bonuses.
17 The described relation remains robust if the analysis is restricted to managers rated ‘fully meets expectations’.
There are potentially competing explanations for our observation. An important question is the direction of causality. One may, for instance, raise the argument that the managers’ performance is affected by their job satisfaction, and that therefore more satisfied managers receive higher bonus percentages. However, the pronounced kink in Germany seems inconsistent with such a reversed causality argument: Without a reference point effect, there is no reason to expect that a manager, who receives a bonus percentage of 100%, exhibits a much larger performance effect when being marginally more satisfied than when being marginally more dissatisfied. Moreover, the US data provide a complementary argument: If satisfaction drove performance and in turn affected bonus percentages, one should also observe such effects within performance ratings, regardless of whether bonus percentages are known or not. However, there is no such relation between percentages and satisfaction in the US, indicating that it is the information about bonus percentages that causally affects satisfaction. We finally note that our conclusion is consistent with both complementary laboratory (as discussed in Section IV) and field experiments.18

Another competing hypothesis for the strong dissatisfaction with bonus percentages below 100% in Germany is that below-average bonuses may potentially signal a lower probability of promotion or of future salary increases, which in turn makes unhappy due to anticipated utility effects. The panel structure of the compensation data allows us to directly test if and how deviations from the 100% bonus are related to the future career progression of a manager. To do this, we estimate a linear probability model regressing the set of remuneration and demographic variables from Model D5 for the year 2005 on a dummy variable equal to one if the manager is promoted for the subsequent business year 2006. The regression includes supervisor fixed effects to estimate the within-department effect of the bonus percentage on the future promotion probability. If the relatively strong dissatisfaction with below-average bonuses is caused by a career signaling effect, we would expect to see a relatively strong negative effect of below-average bonuses on

18 One recent controlled field experiment that independently came to similar conclusions regarding satisfaction is Card et al. (2010), who found that randomly chosen employees of the University of California who learn that they earn a wage below the median for their department and occupation are less satisfied. See Frey and Stephan Meier (2004) and Chen et al. (2010), who also investigate the role of social comparison in controlled field settings.
promotion. However, this is not the case. In fact, managers with above-average bonuses have more reason to be satisfied than managers with below-average bonuses have reason to be dissatisfied: A manager with a ‘fully meets expectations’-rating and 110% bonus has a 8.0% higher probability of being promoted than managers with 100% bonus, whereas the probability of promotion for a manager with 90% bonus decreases only by 5.0% compared to the reference group.\(^{19}\) Furthermore, we get similar results if we use instead a manager’s salary increase as the dependent variable to measure career progression. Each bonus percentage point below 100% in 2005 is associated with 75 Euros less salary increase for the year 2006. However, every percentage point above 100% predicts an estimated salary increase of 131 Euro (see Table A5 in Appendix A). We conclude that the pattern of satisfaction and dissatisfaction with the bonus does not follow the corresponding signaling values of bonuses: the marginal effect of negative deviations from the reference point on promotion and salary increases is small compared to the corresponding effect of positive deviations. That is, career signaling cannot explain the estimated shape of job satisfaction. On the contrary, when anticipatory utility matters, the above findings may actually underestimate the asymmetry in satisfaction due to reference-dependent preferences. Observation 2 summarizes our key finding in this section.

**Observation 2.** Transparent reference point violations lead to decreased job satisfaction. Negative deviations from the reference point have a stronger effect on satisfaction than positive deviations of the same size.

To further illustrate that reference-dependent preferences organize our Observations 1 and 2, we combine a standard approach to reference-dependent social utility (Fehr and Schmidt, 1999), with the reference point being the average bonus, with a framework to analyze subjective performance evaluations (Canice Prendergast and Robert H. Topel, 1999). The results are reported in Table A5 in Appendix A. We observe the same effect with coefficients of similar magnitude when we use a dummy for promotion in the year 2005, or a dummy for promotion in the two-year period from 2005 to 2006 as the dependent variable and include the set of independent variables for 2004. The coefficient of the base salary is negative and significant in all three specifications, indicating that managers on a high level in the company hierarchy are less likely to be promoted further than managers on lower hierarchical levels.
Our model illustrates that, if we take into account the different degrees of transparency regarding the reference point, the patterns observed in our data are a direct implication of combining these approaches. The model is described in detail in Appendix B. Here, we sketch the underlying mechanism.

The managers in our model may be uncertain about others’ bonuses, but they know that bonuses are positively correlated (as is the case in our field context). Each manager has a prior on the distribution of bonuses in mind, and then learns his own bonus payment. The impact of one’s own bonus on expected (social) utility is threefold: First, the bonus may affect material well-being. Second, the bonus payment reveals information about the colleagues’ bonuses. And third, the manager evaluates his own bonus payment relative to his conditional expectation about the average bonus. If the prior is perfectly precise, i.e. the manager knows exactly his relative position, his Fehr-Schmidt type social utility exhibits a kink at the average payoff. If the prior is less precise, the kink becomes less pronounced and the utility function becomes “smooth” around the mean of the prior expectation. In fact, if the precision of the prior goes to zero, the kink in expected social utility disappears entirely, even when the social utility exhibits a pronounced kink. This mirrors our Observation 2.

Endogenizing the assignment of bonuses, we then assume that a supervisor observes a signal on the performance of her managers and has to determine the distribution of a given total bonus budget. Following Prendergast and Topel (1996) and Prendergast (2002), we assume that the supervisor cares for the accuracy of the report but also for the well-being of her managers. It can then be shown that even when performance signals are continuously distributed, bonus payments cluster at the average bonus if the average bonus is transparent: there is a strictly positive probability that managers receive exactly the average bonus. The reason is that when observed performance signals do not differ too strongly, the supervisor prefers to avoid the dissatisfaction caused by reference point violations. If, however, the average bonus and thus one’s relative standing is non-transparent to managers, the optimal distribution of observed bonus payments is continuous, with no specific prominence of the reference point. This prediction is in line with our Observation 1.
III. How do reference point violations affect job performance?

While reference point violations negatively affect employee perceptions, they might be useful to generate performance incentives. In this section, we investigate the competing hypotheses regarding the consequences of reference point violations on performance as outlined in our introduction. Our analysis exploits the fact that we have information about the hierarchical relations between the managers. Hence, we can use a supervisor’s rating as a measure for the performance in her department: 20 If, say, reference point violations negatively affect performance, one would expect that having a higher fraction of managers with a bonus percentage below 100% in one year should lead to a reduced department performance and thus also negatively affect the department supervisor’s performance rating in the subsequent year. 21

Our variable capturing the degree of reference point violation in a department is $DevRef_{st}$, which denotes the share of managers rated below the 100% reference point among all ‘fully meets expectations’-managers evaluated by a supervisor $s$ in a year $t$. By including supervisor fixed effects, we can identify the effect of a within supervisor variation in the frequency of reference point violations on performance in the subsequent year, controlling for time-constant supervisor characteristics and workgroup effects. We estimate the following specification:

$$Rating_{st+1} = \alpha + \beta \cdot DevRef_{st} + \gamma \cdot X_{st} + a_s + \epsilon_{st}.$$  

$X_{st}$ is a vector of individual background variables of a supervisor $s$ and $a_s$ are supervisor fixed effects. The background variables include a supervisor’s age and her hierarchy level in the year $t + 1$. Table 2 reports the results for Germany and the US. All models use a unit normal transformation of the supervisor’s performance rating in $t + 1$ as the dependent variable.

20 This proxy measure for performance has the advantage of being comparable across all departments in the firm.
21 When studying the effect of the bonus on performance, we cannot just conduct our analysis on the level of the managers’ individual performance. Reference point violations might be the result of unobserved factors such as low ability or low willingness to perform, which in turn yields low ratings and low bonuses.
Table 2. Performance effects of reference point violations
(Dependent variable: adjusted supervisor performance rating in subsequent year)

<table>
<thead>
<tr>
<th>Share of Managers with Bonus Percentage &lt;100% among ‘fully meets’ Managers</th>
<th>Germany</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>United States</th>
<th>US1</th>
<th>US2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adequacy x Share of Managers with Bonus Percentage &lt;100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls for Rating Distribution</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>468</td>
<td>468</td>
<td>454</td>
<td>504</td>
<td>504</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-squared (within)</td>
<td>0.05</td>
<td>0.09</td>
<td>0.11</td>
<td>0.04</td>
<td>0.08</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All models are calculated with supervisor fixed effects. Standard errors are given in brackets. *, ** and *** denote significance on the 10%, 5% and 1%-level, respectively. Control variables include the age and hierarchy level of the supervisor. A supervisor’s rating distribution is measured by the share of performance ratings ‘excellent’, ‘above average’ and ‘below average’.

We find that reference point violations negatively affect the performance of a given supervisor in Germany (Model D1). That is, the more ‘fully meets expectations’-managers in a department are pushed below 100% bonus percentage in a given year, the worse is the supervisor’s subsequent rating.

In Model 2, we additionally control for the general evaluation behavior of the supervisor by including the relative shares of the performance ratings different from ‘fully meets expectations’ assigned by a supervisor in t. The benefit of this specification is twofold: For one, while the supervisor fixed effects capture time constant unobserved factors, the heterogeneity of the team regarding individual performance may still vary over time. Controlling for the distribution of ratings captures the effect of this heterogeneity. In addition, the rating distribution determines the restrictions under which a supervisor can (re)distribute the bonus budgets. A supervisor with a team in which many managers have received ‘excellent’ ratings, for instance, may have no choice but to violate reference points of ‘fully meets expectations’ managers, because the system requires that a disproportionally large share of the total budget needs to be spent on the excellent managers. We find that the empirical results are robust: A supervisor in Germany with a
given rating distribution will have a lower performance in the subsequent year when violating the 100% reference point of a higher share of her subordinate managers.

The effects are also economically significant. A rough interpretation of the coefficient is the following: If a supervisor who assigns a bonus percentage of less than 100% to all of her ‘fully meets expectations’ managers instead gave all of them a bonus of 100%, her own performance evaluation in the next year would increase by 0.68 standard deviations. This implies that a supervisor with a performance evaluation at the median would outperform additional 24.6% of all other managers when changing her evaluation behavior in this way.²²

The fact that such reference point violations hurt a supervisor’s rating raises the question why reference points are violated in the first place (although this happens much less than when violations are non-transparent). Our interviews with the company’s HR managers support the view put forward by Prendergast and Topel (1996) and Prendergast (2002) that some supervisors seem to have a preference for accurately assigning bonuses according to performance. A second reason, though, may be out-of-equilibrium behavior: not every supervisor seems to realize the detrimental effect of reference point violations (while some apparently do). Consistent with this view, our research results led the company to redesign the rules by which bonuses are distributed, as well as to improve and extend the training of supervisors assigning those bonuses, suggesting that neither all supervisor behavior nor the institutional design that we study were in equilibrium (see also our concluding section).

²² Alternatively, we estimate Model 2 with a dichotomous dependent variable indicating whether the supervisor is assigned an ‘above average’ rating in the subsequent year. A coefficient for DevRef of -0.253 suggests that a supervisor who violates the reference points of all ‘fully meets expectations’ managers in her department can increase the likelihood of receiving an ‘above average’ rating by more than 25% if she assigns all ‘fully meets expectations’ managers 100% of their budgets. We also note that there is indeed much variation in terms of the frequency of reference point violations in Germany. When supervisors are ranked according to the share of reference point violations in their team, the supervisor at the 33% (66%) percentile gives 40% (80%) of her ‘fully meets expectations’ managers less than the 100% bonus. More detailed analyses, not reported here, indicate that the predominant practice of reference point violation is to push the majority of managers down by a rather small amount, often in order to finance larger bonuses of high performers.
The results of our statistical analysis are strongly supported by the free text comments on the incentive system given by managers at the end of our survey. Many respondents express their concerns that bonus percentages below 100% “de-motivate” managers. For instance one respondent wrote “A good (meets expectations) employee should always be evaluated with 100%, not with 80% or 90%. This always leads to discouragement,” and another one expressed that “If 6 people are evaluated in a department and one gets 105% while the others receive 99%, you have discouraged nearly a complete department”.

Again, the data on the US subsidiary provide the possibility for a falsification exercise. As bonus percentages are not transparent in the US, bonus percentages below 100% should not be as harmful for future performance. Indeed, we do not find any significant effect of reference point violations in the US models.

The regression results are robust in alternative specifications for the German supervisors (see Table A6 in Appendix A). For instance, when using the share of ‘fully meets expectations’ managers below 100% relative to all managers in the department as an alternative measure for the degree of reference point violations (D4), the respective coefficient stays negative and highly significant for Germany. Also, in order to exclude potential explanations driven by changes in the hierarchical composition of the department or changes in team size, we additionally control both for the average bonus budget (D5) and the number of subordinates of a supervisor (D6). The results are also robust when including a lagged dependent variable of the supervisor rating without fixed effects to capture time-varying unobserved characteristics of the supervisor (D7). The coefficient for the frequency of reference point violations remains significantly negative

23 Other examples are “Good managers do not care about the difference of some hundreds of Euros associated with an evaluation with ‘fully meets expectations’ below 100%, but are strongly offended by the evaluation and feel like an underperformer.”; “The absolute bonus amount is not an instrument for motivation. […] A manager who receives a large bonus payment can calculate that his personal target achievement was, for example, 99%. He will immediately realize that he was rated not even as average, and will be discouraged by thousands of Euros.”; “The expectation of every manager who has done a decent job is 100% - every %-point below that is a disappointment.”; “If I assign somebody a bonus below the average, this will not have a positive effect on motivation – no matter how high the payment is.” Comments have been translated from German.

24 For instance, the promotion of a team member to a higher salary grade raises the budget and might possibly lower budgetary pressure, and changes in team size may affect the degrees of freedom in reallocating bonuses.
in all specifications for Germany, while there is no significant effect of reference point violations in any of these specifications for the US (not reported here). We summarize the main result of this section in:

**Observation 3.** Transparent reference point violations subsequently lead to decreased performance.

In search for individual motives underlying the detrimental effect of reference point violations, we find some evidence that negative reciprocal reactions after reference point violations play an important role. Specifically, we can make use of a survey item “My supervisor evaluates my general performance appropriately”, which had to be rated on a 7-point scale from 1 (totally disagree) to 7 (fully agree). The mean value (standard deviation) of the variable is 5.13 (1.48) in Germany. For each supervisor, we calculate the average agreement of her subordinates to this statement (variable “Adequacy”) and include it in the incentive regressions by interacting it with our measure $DevRef$ for reference point violation (see Model D3 in Table 2). The idea is that the more a given reference point violation is perceived as “appropriate”, the less it can be interpreted as unkind, and thus the less it should be prone to a reciprocal response. The regression results strongly support this view. While the coefficient of $DevRef$ remains negative and highly significant, the interaction term is positive, mitigating the negative reference point effect for higher values of perceived adequacy. For the highest possible value (7), the reference point effect is, in fact, entirely counterbalanced. Thus, because reference point violations trigger a negative performance effect only to the extent the subordinates perceive this to be inappropriate, we interpret our result as supporting the notion that reciprocity is an important driver for the performance effect. We provide more evidence for the importance of reciprocity in a laboratory experiment, which we report in the next section.25

---

25 Our data do not imply that differentiation between managers per se hampers performance: e.g., some of our analyses – not reported here – seem to indicate that differentiating in ratings might improve performance. However, controlling for rating, differentiation in bonus percentages is always associated with a subsequent drop in performance.
IV. A laboratory experiment

This section provides another robustness check for our observations, and sheds further light on the underlying behavioral mechanisms based on a laboratory study. The laboratory environment shares the relevant features of our company’s bonus system, such as a fixed sum of bonus payments and the differences in the transparency of bonus percentages. At the same time, it abstracts away from various potentially complicating factors in the field context, such as signaling issues and cultural differences.26

In the experiment, two anonymous, randomly matched participants worked on a joint essay sitting in separate cabins using GoogleDocs, a web-based word processor (http://google.docs.com/). Each group had to search for and describe web-pages about electric cars, and then list arguments in favor of and against the use of electric cars (see Appendix C for instructions). How the task was approached and the work was divided could be freely negotiated between the subjects via an online chat. A third participant, acting as the supervisor, had then to rank workers according to their performance based on excerpts from the respective contributions to the essay.27 Supervisors and workers interacted anonymously.

We implemented a 2x2 design to systematically analyze the interaction of bonus inequality (EQUAL and UNEQUAL) and transparency (TRANSPARENCY and NON-
TRANSPARENCY) on satisfaction and behavior. In all treatments participants were initially unaware of the magnitude of the bonus budget. In the UNEQUAL treatments, the high-ranked worker received 120% (24 Euros) of the average bonus budget, while the low-ranked worker received 80% (16 Euros). In the EQUAL treatments, both the high- and the low-ranked worker received 100% of the average budget (20 Euros). In the TRANSPARENCY treatments, participants were informed about their absolute bonuses and their relative standing, i.e. their percentage of the average bonus budget. In the NON-TRANSPARENCY treatments, participants were only informed about their absolute bonus payments.

After bonus payments were assigned, we asked each worker for two additional choices (workers knew that the experiment had three parts, but were not told the nature of the latter parts before the respective earlier parts were finished). First, the same two workers who participated in the first part were matched to play a standard laboratory public good game. Next, each of the workers received an extra endowment of 2 Euros and played a dictator game: they could transfer any part of the endowment to the supervisor. The transferred amount was then doubled to depict efficiency gains for the supervisor from the workers’ ‘efforts’. Participants did not receive feedback about their payoffs from the two additional parts before the end of the experiment. As we explain below, the two additional choices may reveal the underlying motives to (not) exert effort in response to reference point violations.

We conducted four experimental sessions in November 2010 in the Cologne Laboratory of Economic Research with altogether 120 subjects, yielding 10 independent observations for each treatment. Participants were recruited using the online recruitment system ORSEE (Ben Greiner, 2004). The chat and the second and third part of the experiment were implemented using z-tree (Fischbacher, 2007). Every session lasted about two hours; the average payoff (standard deviation) was 28.02 Euros (6.50 Euros) including a show-up fee of 4 Euros for workers and 18 Euros for supervisors. After each session, participants filled out a post-experimental questionnaire, in which, among other things,

28 Both workers received an endowment of 5 Euros which they could contribute to a joint project. The sum of contributions was multiplied by the factor 1.5 and equally distributed among the workers.
they were asked about their satisfaction with bonus payments measured on the same 1-to-7 scale as in our manager survey. Participants then privately received their payments and left the laboratory.

From both, the field observation and our model, we hypothesized that transparency amplifies the dissatisfaction from reference point violations. As Figure 3 illustrates, our experiment supports the hypothesis. There are no significant differences in reported satisfaction between subjects with different bonus percentages in the NON-TRANSPARENCY treatment (all tests yield \( p \)-values > 0.1). In the transparent condition, however, the relevant differences are statistically significant.\(^\text{29}\) Moreover, negative deviations from 100% bonus have a substantially stronger impact on satisfaction than positive deviations: workers with 80% bonus report a satisfaction score that is 2.4 points lower than subjects with 100%, while receiving 120% increases satisfaction by only 0.9 points.

Figure 3. Experimental results (averages per treatment)

We now study how the bonus from the first part of the experiment affects the workers’ willingness to cooperate in the second and third part, as well as the underlying

\(^\text{29}\) Comparing subjects with 80% and 100% bonus and subjects with 100% and 120% bonus using two-sided Mann-Whitney-U tests, and subjects with 80% and 120% bonus using two-sided Sign tests yields significance levels of \( p = 0.002 \), \( p = 0.003 \) and \( p = 0.002 \), respectively.
motivations. More specifically, if workers are motivated to reduce payoff inequality among themselves after a transparent reference point violation, we would expect to see that below-average paid workers contribute less in the subsequent public goods game in the second part of the experiment. The second panel in Figure 3 illustrates that this is not the case.\textsuperscript{30} However, if workers are motivated to reciprocally punish the supervisor, we would instead expect to see that the pattern of dictator giving mirrors the workers’ dissatisfaction with the bonus in the third part of the experiment. The data indeed confirm this latter hypothesis. If relative standings are made transparent, the spread between transfers of 80%- and 100%-subjects is twice as large as in the NON-TRANSPARENCY treatment. Below-average paid workers transfer 7.5% of the endowment versus 27.5% of average-paid workers ($p = 0.014$, two-sided Mann-Whitney-U test) and 22.5% of above-average paid workers ($p = 0.028$, two-sided Wilcoxon-Matched-Pairs-Signed-Ranks test); workers with 100% and 120% do not differ significantly in their transfers ($p = 0.817$, two-sided Mann-Whitney-U test). Finally, as in the case of pay satisfaction, we do not observe a statistically significant difference between any worker group in the NON-TRANSPARENCY treatment.\textsuperscript{31} That is, only transparent reference point violations induce a negatively reciprocal response towards the supervisor.

Because the laboratory environment controls for various institutional, social and strategic complications in the field, it serves as another robustness test for our field observations. Overall, we find that our field and experimental study, together with the model, provide a remarkably coherent picture of the impact of reference point violations. In particular, the transparency of reference point violations turns out to be a key factor for the evaluation and effectiveness of bonus payments: falling behind others reduces satisfaction and subsequently hampers performance and cooperation. At the same time, our experiment

\textsuperscript{30} Interestingly, subjects with 80% bonus in the transparent condition even tend to contribute more to the public good than subjects with 100% bonus ($p = 0.075$, two-sided Mann-Whitney-U test). One might speculate that some of these subjects attempt to catch up with the workers’ ‘overall average effort’ after one’s contribution to the team production in the first part of the experiment has been evaluated below average. All other comparisons are insignificant on conventional levels.

\textsuperscript{31} Dictator transfers increase monotonically with assigned bonus percentages (workers with 80%, 100% and 120% send 14.5%, 24.4% and 31.3% of their respective endowments to the supervisors), but the differences are not significant: using two-sided Wilcoxon-Matched-Pairs-Signed-Ranks tests (80% versus 120%) and two-sided Mann-Whitney-U tests (80% versus 100% and 100% versus 120%) yields $p$-values $> 0.1$. 

27
provides complementary insights regarding the role of reciprocity in explaining the group performance effect.32

V. Conclusion

We investigate how bonus payments affect satisfaction and performance of managers in a large, multinational company. The company’s short term incentive system produces a clear, naturally occurring reference point for bonus comparisons. Utilizing a unique data set that combines survey and compensation data, we find that if a manager’s bonus falls behind the reference point, self-reported job satisfaction is significantly reduced, while being ahead does hardly affect a manager’s satisfaction. Moreover, reference point violations reduce subsequent performance: A supervisor who pushes more managers in her department below the reference point subsequently gets a lower performance rating herself. One implication of the harmful effects of reference point violations is that bonus payments are strongly compressed towards the reference point. These effects are mitigated, though, if information about one’s relative standing towards the reference point is withheld.

Our findings are robust. For one, our data allowed us to rule out several competing explanations for the impact of transparent reference point violations, such as signaling.33 Also, we show that the role of transparency follows straightforwardly from widely used

32 The experimental finding is consistent with recent and mostly independent experimental work on reciprocity. Regarding laboratory research, Gächter and Christian Thöni (2010) conducted several laboratory gift-exchange games and found that effort reductions in response to disadvantageous relative wage positions were triggered only if wages were set intentionally by a principal, supporting the notion that reference point violations can trigger negative reciprocity. There is also recent evidence from laboratory and field experiments that disadvantageous relative positions negatively affect effort exertion and working performance (see, for example, Clark, David Masclet and Marie-Claire Villeval, 2010; Alain Cohn et al., 2010; Greiner, Ockenfels and Peter Werner, forthcoming). On the theory side, we mention that in Falk and Fischbacher’s (2006) reciprocity model, the kindness of the opponent is judged with respect to the chosen (in)equality of outcomes. If this is the case, inequality is not only dissatisfying but also the trigger for reciprocal reactions. This interpretation of the data is in line with our observations.

33 One reason for why we find quite strong and robust effects of transparent reference point violations may be that the 100% bonus percentage is both a reasonable social reference point and a reasonable expectation for ‘fully meets expectations’ managers. While we have some evidence that social comparison is at least part of the explanation, it is an interesting task for future research to further disentangle the contribution of those two potential reference points.
behavioral economics approaches when we introduce the possibility that managers are uncertain about their relative standing. Moreover, adding a framework of subjective performance evaluations to our model can explain that bonuses are clustered at the reference point if reference point violations are transparent. Finally, our main observations are supported by a complementary laboratory experiment. The lab study mirrors the relevant features of the field incentive system, controls for potentially confounding influences, and provides further insights about the underlying behavioral mechanisms behind the performance effect.

To us, one of the main goals of this line of research is not only to develop a better understanding of the actual effectiveness of bonus systems, but also to be increasingly able to make recommendations for better systems in practice. Survey studies indicate that the demand for better systems is large: The consultancy firm Towers Perrin, for instance, stated in its report of a large company survey conducted in 2007 that “more than three-quarters of responding organizations have changed their variable pay programs in the past three years and nearly half expect to implement more changes in the future”. However, the effectiveness of many commonly used incentive instruments, and their interaction, is not well understood yet.

For instance, while our study shows that transparency comes at a cost due to reference-dependent preferences, our findings do not necessarily imply that transparency is per se harmful. A lack of transparency may also have negative consequences, related to procedural fairness perceptions (see, in psychology, John Schaubroeck, Douglas R. May, and F. William Brown, 2000, and Steve Werner and Deniz S. Ones, 2000, and in

---

34 Our observations regarding transparency are also consistent with the social psychology literature, where a series of classic findings demonstrate that social judgments critically depend on which comparison standards are made accessible in the judgmental situation. That is, comparison standards that are not (made) accessible are not used (see Thomas Mussweiler and Lysann Damisch, 2008, for a review of the more recent relevant cognitive psychology research, and Chen et al., 2010, and Daniel J. Benjamin, James J. Choi, and A. Joshua Strickland, 2010, for an application in the economics literature).

35 The report refers to the 2007 ‘Towers Perrin rewards challenges and changes survey’ including 637 HR and compensation executives from 21 countries. The ongoing financial crisis has triggered an even more forceful debate about the optimal structure of bonus plans.

36 A related observation has been made by Frank and Cass R. Sunstein (2001, p.343), who noted that at the University of Chicago Law School “there is an exceptionally strong norm against public discussion of salaries, even among good friends. The evident basis of the norm is to prevent dissatisfaction about relative position in the face of satisfaction with absolute position.”
economics, Bolton, Brandts and Ockenfels, 2005, and Frey and Stutzer, 2005). One reason is that a lack of transparency bears the risk of inequitable pay rates that would be objected if transparent. Our study for instance shows that a less transparent system may exhibit significantly more reference point violations than a more transparent system.\footnote{Transparency of pay rates and bonuses is also on the political agenda in the US. For instance, in August 2009, Colorado joined California, Michigan and Illinois in protecting ‘employees’ right to discuss their wages’. The so-called “Wage Transparency Act” prohibits employers from taking adverse actions against employees who discuss their wages with others.} Because the role of transparency is not yet well investigated in (behavioral) economics, we are not sufficiently confident to claim that transparency about relative positions is generally not recommendable. Yet we think it is safe to conclude from our study that limiting the scope for bonus and pay comparisons is often beneficial. In fact, the bonus system studied in this paper has been recently changed, as a reaction to our study. One important change was to fix the bonus percentage for each performance grade in order to avoid detrimental within-grade comparisons without reducing the overall transparency.

Finally, we conclude on a more general level that a good understanding of how reference points for bonus comparisons evolve and how reference point formation can be influenced, appears to be of critical importance both for effective practical incentive design and for the further development of the theory of reference-dependent preferences. Our study is one step in this direction.
References


Gächter, Simon, and Ernst Fehr. 2002. “Fairness in the Labour Market: a Survey of Experimental Results.” In Surveys in Experimental Economics. Bargaining,


Appendices for online publication only.

Appendix A: Further descriptive statistics and robustness checks

Table A1. Distribution of performance ratings (in % of all managers)

<table>
<thead>
<tr>
<th></th>
<th>Germany 2006</th>
<th>US 2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘excellent’</td>
<td>0.6</td>
<td>1.7</td>
</tr>
<tr>
<td>‘above average’</td>
<td>22.4</td>
<td>32.8</td>
</tr>
<tr>
<td>‘fully meets expectations’</td>
<td>74.1</td>
<td>64.0</td>
</tr>
<tr>
<td>‘below average’</td>
<td>2.8</td>
<td>1.5</td>
</tr>
<tr>
<td>‘inadequate’</td>
<td>0.1</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table A2. Distribution of agreements to the statement: I am very satisfied with my job.
(in % of all managers who participated in the survey)

<table>
<thead>
<tr>
<th></th>
<th>Germany 2006</th>
<th>US 2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - totally disagree</td>
<td>1.0</td>
<td>1.1</td>
</tr>
<tr>
<td>2</td>
<td>3.0</td>
<td>3.9</td>
</tr>
<tr>
<td>3</td>
<td>6.0</td>
<td>6.6</td>
</tr>
<tr>
<td>4</td>
<td>10.2</td>
<td>13.9</td>
</tr>
<tr>
<td>5</td>
<td>27.9</td>
<td>27.5</td>
</tr>
<tr>
<td>6</td>
<td>40.0</td>
<td>34.4</td>
</tr>
<tr>
<td>7 - fully agree</td>
<td>11.9</td>
<td>12.7</td>
</tr>
<tr>
<td>Mean</td>
<td>5.29</td>
<td>5.24</td>
</tr>
<tr>
<td>Std.Dev.</td>
<td>1.25</td>
<td>1.25</td>
</tr>
</tbody>
</table>
Table A3. Determinants of job satisfaction in Germany – robustness checks  
(Ordered Probit models with the ‘job satisfaction’ score as dependent variable)

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Job Satisfaction (Score ‘7’ excluded)</th>
<th>Job Satisfaction (Score ‘7’ excluded)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ordered Probit</td>
<td>Ordered Probit</td>
</tr>
<tr>
<td>Salary (000s)</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>[0.004]</td>
<td>[0.004]</td>
</tr>
<tr>
<td>Bonus Payment (000s)</td>
<td>0.000</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>[0.009]</td>
<td>[0.011]</td>
</tr>
<tr>
<td>Dummy ‘Above meets’</td>
<td>0.282***</td>
<td>[0.103]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dummy ‘Below meets’</td>
<td>-0.438*</td>
<td>[0.227]</td>
</tr>
<tr>
<td>Positive Deviation from Bonus</td>
<td>-0.009</td>
<td>-0.008</td>
</tr>
<tr>
<td>Budget (=100%)</td>
<td>[0.006]</td>
<td>[0.014]</td>
</tr>
<tr>
<td>Negative Deviation from Bonus</td>
<td>-0.025***</td>
<td>-0.030***</td>
</tr>
<tr>
<td>Budget (=100%)</td>
<td>[0.006]</td>
<td>[0.007]</td>
</tr>
<tr>
<td>Sample</td>
<td>All</td>
<td>‘fully meets’</td>
</tr>
<tr>
<td>Observations</td>
<td>1828</td>
<td>1367</td>
</tr>
<tr>
<td>Log-Likelihood</td>
<td>-2391</td>
<td>-1818</td>
</tr>
</tbody>
</table>

Standard errors are given in brackets. * and *** denote significance on the 10% and 1%-level, respectively. Control variables include age interval dummies, gender, total years of affiliation to the company, a dummy for a promotion in the previous year, and dummies for company units.
Table A4. Determinants of job satisfaction – robustness checks (linear models with unit-normal transformation of the ‘job satisfaction’ score as dependent variable)

<table>
<thead>
<tr>
<th>Country</th>
<th>Model</th>
<th>Germany</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salary (000s)</td>
<td>-0.003</td>
<td>-0.006</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.004]</td>
<td>[0.004]</td>
<td></td>
</tr>
<tr>
<td>Bonus Payment (000s)</td>
<td>0.012</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.010]</td>
<td>[0.006]</td>
<td></td>
</tr>
<tr>
<td>Dummy ‘Above meets’</td>
<td>-0.037</td>
<td>0.094</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.129]</td>
<td>[0.196]</td>
<td></td>
</tr>
<tr>
<td>Dummy ‘Below meets’</td>
<td>-0.613***</td>
<td>-0.927*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.206]</td>
<td>[0.520]</td>
<td></td>
</tr>
<tr>
<td>z ≥ 110%</td>
<td>0.061</td>
<td>0.144</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.151]</td>
<td>[0.319]</td>
<td></td>
</tr>
<tr>
<td>100% &lt; z &lt; 110%</td>
<td>0.013</td>
<td>0.065</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.093]</td>
<td>[0.281]</td>
<td></td>
</tr>
<tr>
<td>90% ≤ z &lt; 100%</td>
<td>-0.182**</td>
<td>-0.065</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.089]</td>
<td>[0.272]</td>
<td></td>
</tr>
<tr>
<td>z &lt; 90%</td>
<td>-0.295**</td>
<td>-0.112</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.139]</td>
<td>[0.346]</td>
<td></td>
</tr>
<tr>
<td>Sample</td>
<td>All</td>
<td>All</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>2091</td>
<td>956</td>
<td></td>
</tr>
<tr>
<td>R-squared (within)</td>
<td>0.04</td>
<td>0.04</td>
<td></td>
</tr>
</tbody>
</table>

The dependent variable is a unit normal transformation of the job satisfaction score. \( z \) denotes a manager’s bonus percentage. All models are calculated with supervisor fixed effects. Standard errors are given in brackets. *, ** and *** denote significance on the 10%, 5% and 1%-level, respectively. Control variables include age interval dummies, gender, total years of affiliation to the company and a dummy for promotion in the last year.
Table A5. Bonus percentages and career progression in Germany

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>FE</td>
<td>FE</td>
<td>FE</td>
<td>FE</td>
<td>FE</td>
<td>FE</td>
</tr>
</tbody>
</table>

Independent variables from year

<table>
<thead>
<tr>
<th>Salary (000s)</th>
<th>-0.002***</th>
<th>-0.003***</th>
<th>-0.005***</th>
<th>0.000</th>
<th>0.003</th>
<th>-0.002</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[0.001]</td>
<td>[0.001]</td>
<td>[0.004]</td>
<td>[0.006]</td>
<td>[0.010]</td>
<td></td>
</tr>
<tr>
<td>Bonus Payment (000s)</td>
<td>0.000</td>
<td>0.000</td>
<td>-0.002</td>
<td>0.016</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>[0.002]</td>
<td>[0.003]</td>
<td>[0.004]</td>
<td>[0.011]</td>
<td>[0.017]</td>
<td>[0.026]</td>
</tr>
<tr>
<td>Positive Deviation from Bonus Budget (=100%)</td>
<td>0.008**</td>
<td>0.008**</td>
<td>0.013***</td>
<td>0.131***</td>
<td>0.112***</td>
<td>0.180***</td>
</tr>
<tr>
<td></td>
<td>[0.003]</td>
<td>[0.003]</td>
<td>[0.004]</td>
<td>[0.018]</td>
<td>[0.018]</td>
<td>[0.029]</td>
</tr>
<tr>
<td>Negative Deviation from Bonus Budget (=100%)</td>
<td>-0.005***</td>
<td>-0.006***</td>
<td>-0.010***</td>
<td>-0.075***</td>
<td>-0.063***</td>
<td>-0.119***</td>
</tr>
<tr>
<td></td>
<td>[0.002]</td>
<td>[0.002]</td>
<td>[0.004]</td>
<td>[0.009]</td>
<td>[0.010]</td>
<td>[0.017]</td>
</tr>
</tbody>
</table>

Sample: ‘fully meets’ ‘fully meets’ ‘fully meets’ ‘fully meets’ ‘fully meets’ ‘fully meets’

Observations: 2454 1922 1922 2447 1914 1926

R-squared (within): 0.04 0.06 0.13 0.19 0.16 0.23

‘Promotion’ is a dummy variable equal to one if a manager is on a higher hierarchy level in a given year than in the base year of the model. Salary increases are measured in 1,000s of Euros. All models are calculated with supervisor fixed effects. Standard errors are given in brackets. ** and *** denote significance on the 5% and 1%-level, respectively. Control variables include age interval dummies, gender and total years of affiliation to the company.
Table A6. Performance effects of reference point violations in Germany – robustness checks (dependent variable: adjusted supervisor performance rating in subsequent year)

<table>
<thead>
<tr>
<th>Model</th>
<th>D4</th>
<th>D5</th>
<th>D6</th>
<th>D7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FE</td>
<td>FE</td>
<td>FE</td>
<td>OLS</td>
</tr>
<tr>
<td>Share of Managers rated ‘fully meets’ with Bonus Percentage &lt;100% in department</td>
<td>-0.786***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.297]</td>
</tr>
<tr>
<td>Share of Managers with Bonus Percentage &lt;100% among ‘fully meets’ Managers</td>
<td>-0.689***</td>
<td>-0.671***</td>
<td>-0.246*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.217]</td>
<td>[0.221]</td>
<td>[0.128]</td>
<td></td>
</tr>
<tr>
<td>Bonus Budget per Manager</td>
<td>-0.014</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.020]</td>
</tr>
<tr>
<td>Team Size</td>
<td></td>
<td></td>
<td></td>
<td>0.685***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.093]</td>
</tr>
<tr>
<td>Controls for Rating Distribution</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>468</td>
<td>468</td>
<td>468</td>
<td>468</td>
</tr>
<tr>
<td>R-squared (within)</td>
<td>0.07</td>
<td>0.09</td>
<td>0.09</td>
<td>0.23¹</td>
</tr>
</tbody>
</table>

Models D4, D5, D6 are calculated with supervisor fixed effects. Standard errors are given in brackets. In Model D7, robust standard errors clustered on the level of supervisors are reported. * and *** denote significance on the 10% and 1%-level, respectively. All models include control variables for the age and hierarchy level of the supervisor. Model D7 additionally includes dummy variables for company entities. A supervisor’s rating distribution is measured by the share of performance ratings ‘excellent’, ‘above average’ and ‘below average’.

¹ Overall R-squared value
Appendix B: A simple theoretical framework

Motivated by Observations 1 and 2 in the main text, this appendix presents a simple theoretical framework to fix ideas about the impact of transparency on managers’ social utility (B.1) and supervisors’ evaluation behavior (B.2).

B.1 Transparency and social comparison

We assume that each manager has a Fehr and Schmidt (1999) type piecewise linear utility function, comparing his own bonus \( b_i \) to the average bonus \( B \) paid to his colleagues:

\[
\begin{align*}
\text{if } & b_i \geq B : u(b_i, B) = \eta \cdot b_i - \beta (b_i - B) \\
\text{if } & b_i < B : u(b_i, B) = \eta \cdot b_i - \alpha (B - b_i)
\end{align*}
\]

where \( \eta \) measures the weight placed on the absolute bonus, \( \alpha \) measures the manager’s dislike of disadvantageous and \( \beta \) (< \( \alpha \)) that of advantageous inequality. The model implies that reference point violations cause a marginal utility loss \( \alpha \) in addition to \( \eta \) that comes with the corresponding loss of absolute payoffs. For simplicity, assume that there is a continuum of managers. In order to capture the effect of transparency, we assume that a manager is uncertain about the social reference point. Managers share a prior belief that the mean bonus \( B \) is drawn from a normal distribution \( B \sim N(m, \sigma^2) \). Furthermore, each manager knows that his own bonus \( b_i \) is equal to \( b_i = \epsilon + B \) where \( \epsilon \sim N(0, \sigma^2) \).

In this framework, \( \sigma^2 \) is a straightforward measure of transparency. If \( \sigma^2 \) is equal to zero, a manager knows exactly where he stands relative to his colleagues (as it is the case in the German branch of the studied company). If, however, \( \sigma^2 \) is positive, there is uncertainty about one’s relative standing (as it is the case in the US branch). A manager’s posterior belief on the reference point is then:

\[
E[b | \epsilon] = \frac{\sigma^2}{\sigma^2 + \sigma^2} m + \frac{\sigma^2}{\sigma^2 + \sigma^2} b \quad \text{and} \quad V[b | \epsilon] = \frac{\sigma^2 \sigma^2}{\sigma^2 + \sigma^2}.
\]
We now compute a manager’s conditional expected utility, after he has learned his own bonus payment, as:

\[
E[u(b, B)|b] = E[I_{B>0}((\eta - \beta)b + \beta B)|b] + E[I_{B>0}((\eta + \alpha)b - \alpha B)|b]
\]

\[
= (\eta + \alpha)b - \Pr(B \leq b|b)\beta(b(\alpha + \beta) + \beta E[I_{B>0}B|b] - \alpha E[I_{B>0}B|b])
\]

After applying standard results on the truncated normal distributions and some rearrangements, we obtain the following:

**Proposition B1:** The manager’s expected utility is given by

\[
E[u(b, B)|b] = \eta b + \alpha \frac{\sigma^2}{\sigma^2 + \sigma_e^2}(b - m)
\]

\[
+ (\alpha + \beta) \left( \frac{\sigma_e^2(m - h)}{\sigma^2 + \sigma_e^2} \Phi \left( \frac{\sigma_e(b - m)}{\sqrt{\sigma^2 + \sigma_e^2}} \right) - \Phi \left( \frac{\sqrt{\sigma^2 + \sigma_e^2} \eta}{\sqrt{\sigma^2 + \sigma_e^2}} \right) \right)
\]

(2)

The dissatisfaction from a reference point violation (and the corresponding kink) disappears as the prior becomes uninformed \((\sigma^2 \to \infty)\), even when managers are strictly inequity averse.

In case of full transparency \((\sigma^2 = 0)\), the marginal utility loss due to falling behind others is \(\alpha > 0\) (Equation (1)). Because the corresponding gain \(\beta\) from advantageous inequality is smaller, there is a ‘kink’ at the reference point. If, however, the prior becomes uninformed \((\sigma^2 \to \infty)\), the conditional expected utility (Equation (2)) converges to \(E[u(b, B)|b] = \eta b - (\alpha + \beta) \frac{\sigma_e^2}{\sigma_e^2 + \sigma_e^2}\), which is independent of the reference point. The reason is that a manager’s conditional expectation on the other managers’ bonus payments converges to his own bonus payment \(b\) \((= \lim_{\sigma \to \infty} E[B|b])\). As a result, there can be no loss in expected utility due to reference point violations.

For intermediate values of reference point transparency \(\sigma^2\), it is instructive to study the shape of the conditional expected utility graphically. The solid curve in Figure B1 shows the conditional utility function when there is full transparency (Equation (1)). The dashed curves show the conditional expected utility for different values of \(\sigma^2\).\(^{38}\) The grey line

---

\(^{38}\) The variance \(\sigma^2\) is equal to 4, 8 and 12, respectively. The other parameter values are \(m = 100, \sigma_e = 5, \alpha = 0.6, \beta = 0.2,\) and \(\eta = 0.3\). If the managers’ utility is not affected by their absolute but only by their relative
finally shows the expected utility function for $\sigma^2 \rightarrow \infty$. The figure demonstrates that decreasing transparency reduces dissatisfaction from reference point violations and ‘softens’ the kink.

Figure B1. Conditional expected utility as a function of the bonus $b$

Our model can be thought of as generalizations of two related approaches. For one, we generalize an aspect of a model proposed by Card et al. (2010), in which individuals care for their absolute as well as their relative pay as compared to a reference group, and may either receive information on their relative pay or not. When there is no information about the others’ bonuses, Card et al. assume that the conditional expectation about the average wage in the reference group is equal to the individual’s own wage, which is endogenized in our model as $\sigma^2 \rightarrow \infty$. Second, our model is consistent with findings in the social psychology literature that demonstrate that social judgments critically depend on which comparison standards are made accessible in the judgmental situation. That is, comparison standards that are not (made) transparent are not used (see references in Section V).

bonuses (i.e. $\eta = 0$), the expected utility becomes entirely flat for an uninformed prior even when there is a pronounced kink with complete information.
B.2 Reference-dependent preferences and the allocation of bonuses

In a second step we analyze the impact of reference-dependent preferences on performance evaluation. For this, we take a standard model of subjective performance evaluations, as developed by Prendergast and Topel (1996) and Prendergast (2002), and only add inequality aversion to the managers’ preferences as specified above.

Following Prendergast and Topel (1996) and Prendergast (2002), we assume that the supervisor cares for the well-being of her subordinates and for the accuracy of the bonus. More specifically, we assume that the supervisor’s utility function is linearly increasing in the subordinates’ utility, and that the supervisor incurs a utility loss $\nu \cdot (s_i - b_i)^2$ when assigning the bonus $b_i$ to a manager $i$ with true performance $s_i$. Applying this framework to our context, we also assume that the supervisor has to allocate an average bonus $B$, and – for simplicity – that she only has two subordinates, 1 and 2. We start with a fully transparent social reference point. The supervisor’s utility function then is

$$U(h) = \mu u_1(h_1, b_2) + \mu u_2(h_2, b_1) - \nu (s_1 - h_1)^2 - \nu (s_2 - h_2)^2,$$

where $u_i(h_i, b_i)$ is given by Equation (1). Substituting the budget constraint $b_1 + b_2 = 2B$ and the managers’ utility functions, the supervisor’s decision problem is to maximize

$$U(h) = \mu (\eta 2B - 2(\alpha + \beta) \cdot |h_1 - B|) - \nu (s_1 - h_1)^2 - \nu (s_2 - 2B + h_1)^2.$$

This function is continuous but not continuously differentiable as it has a kink at $h_1 = B$. The second derivative is negative to the left and to the right of $B$ and $\partial U_1(B) < \partial U_1(B)$, and hence, the function is strictly concave. Suppose that the optimal bonus $b_i^*$ is strictly smaller than $B$. In that case it must be characterized by the first order condition

$$2\mu(\alpha + \beta) + 2\nu(s_1 - b_1^*) - 2\nu(s_2 - 2B + b_1^*) = 0$$

implying

$$b_1^* = B + \frac{\mu(\alpha + \beta)}{2\nu} + \frac{s_1 - s_2}{2}.$$
But the first order condition only characterizes the optimal choice if at this point indeed $b^*_1 < B$, which is the case if $s_1 - s_2 < -\frac{\mu(a+\beta)}{v}$. By symmetry we thus obtain the supervisor’s optimal strategy for fully transparent settings:

$$b^*_1(s_1, s_2) = \begin{cases} 
B + \frac{s_1 - s_2 + \mu(a+\beta)}{2v} & \text{if } s_1 - s_2 < -\frac{\mu(a+\beta)}{v} \\
B & \text{if } -\frac{\mu(a+\beta)}{v} \leq s_1 - s_2 < \frac{\mu(a+\beta)}{v} \\
B + \frac{s_1 - s_2 - \mu(a+\beta)}{2v} & \text{if } s_1 - s_2 \geq \frac{\mu(a+\beta)}{v}
\end{cases} \quad (3)
$$

In a fully non-transparent setting, however, the kink in the managers’ utility function disappears, as stated in Proposition B1. Hence, the bonus payment becomes

$$b^*_1(s_1, s_2) = B + \frac{s_1 - s_2}{2}.
$$

Thus, compared to a non-transparent setting (as in the US in our field context), transparency about the budget leads to compressed performance evaluations (as we see in Observation 1),\(^{39}\) as there is an upward bias for the manager with the lower and a downward bias for the manager with the higher performance. Moreover, even when there is an atomless continuous distribution of the true performance levels, our model predicts a mass point of evaluations at $B$. To illustrate this, assume that $s_1$ and $s_2$ are $iid$ and drawn from a normal distribution with variance $2\sigma^2$. Hence, the bonuses awarded without transparency about the budget $b^T = B + \frac{s_1 - s_2}{2}$ are also normally distributed, with mean $B$ and variance $\sigma^2$. But when the budget is transparent, the bonuses $b^T$ will not be normally distributed as the following result shows:

**Proposition B2:** Transparency leads to less reference point violations. Specifically, assuming that performance is normally distributed, bonuses are also normally distributed in a non-transparent system, but have a mass point at the reference point $B$ in a transparent system.

\(^{39}\) This gives a rationale for the so-called ‘centrality bias’ in subjective performance evaluations (see for instance Prendergast and Topel, 1993).
Proof: We have \( b^T = b' + \frac{\mu(a + \beta)}{2v} \) if \( b' < B - \frac{\mu(a + \beta)}{2v} \), and \( b^T = b' - \frac{\mu(a + \beta)}{2v} \) if \( b' \geq B + \frac{\mu(a + \beta)}{2v} \). Hence, for the lower tail of the distribution of \( b^T \) we must have that

\[
\Pr(b^T < b) = \Pr(b' + \frac{\mu(a + \beta)}{2v} < b) = \Phi\left(\frac{1}{\sigma}(b - \frac{\mu(a + \beta)}{2v} - B)\right)
\]
as long as \( b' < B - \frac{\mu(a + \beta)}{2v} \) or \( b^T = b' + \frac{\mu(a + \beta)}{2v} < B \). By symmetry this also holds for the upper tail. Furthermore the probability mass at \( B \) is equal to

\[
\Pr\left(B - \frac{\mu(a + \beta)}{2v} \leq b' < B + \frac{\mu(a + \beta)}{2v}\right) = \Phi\left(\frac{\mu(a + \beta)}{2\sigma}\right) - \Phi\left(-\frac{\mu(a + \beta)}{2\sigma}\right) = 2\Phi\left(\frac{\mu(a + \beta)}{2\sigma}\right) - 1
\]
which is strictly larger than zero as long as \( \alpha + \beta > 0 \), as well as strictly increasing in \( \alpha \) and \( \beta \). \( \blacksquare \)
Appendix C: Experimental Instructions (Translation from German)

C.1 Instruction for Workers

Instructions: General Information

Welcome to the experiment! From now on, please do not communicate with other participants. If you have a question, please raise your hand! We will come to you and answer your question. If you violate these rules, we have to exclude you from the experiment and all payoffs.

In this experiment you can earn money. How much depends on your decisions and the decisions of other participants. We use ECU (Experimental Currency Unit) as the laboratory currency. At the end of the experiment, your payoff in ECU is converted into Euro and paid out in cash. The exchange rate is 20 ECU = 1 Euro.

The experiment consists of three parts. After the experiment, you receive the sum of payoffs from these parts. In addition, you receive 4 Euros for your participation in the experiment, which is paid out at the end regardless of the decisions.

Instructions: First Part

In the first part of the experiment, you will be matched with another participant. During the next 30 minutes, you will jointly work on the following task:

*Which are the best and the worst homepages on electric cars you can find? Why? List the best arguments for and against electric cars.*

You and the other participant have to prepare a joint document for the summary. In the document, contributions of each participant have to be marked. The preparation of the document will be explained on the following pages.

You and the other participant are assigned to an evaluator. After your document has been prepared, the evaluator receives an excerpt of both participants’ contributions and evaluates the performance of each participant.

The evaluators are other participants in the experiment who receive a fixed payment for the evaluation of the excerpts.

You and the other participant will receive an individual bonus payment depending on the evaluations. There is a budget for bonus payments; on average there is a certain bonus amount available for each participant.

[Treatment TRANSPARENCY] Individual bonus payments can be as low as 80% and as high as 120% of this average amount.

After the first part of the experiment, you will be informed about your bonus payment.

The identity of the other participant is confidential, and no other participant will be informed about your identity: your decisions are anonymous.

Description of the task

During the next 30 minutes, you and the other participant will work on the following task: *Which are the best and the worst homepages on electric cars you can find? Why? List the best argument for and against electric cars.*

You will work on the task online on the “Google Docs” website. Please open the Mozilla Firefox browser by clicking the symbol on your desktop. In the Favorites menu (see Screenshot 1 below), you find the link for your joint document.
The link “Task Group X”, will direct you to the joint document. The following screen appears (Screenshot 2), in which you can enter the text. You and the other participant work simultaneously in this document. You will immediately see the other participant’s entries; likewise, the other participant will see your entries in real-time.

The text has to be written in German. Copied parts have to be identified. Please leave the header of the document unchanged and enter your text in the designated part (see Screenshot 2). Please list all references you used (it is sufficient to paste the hyperlinks of the references in the document).

“Google Docs” will automatically save your entries. If you close the browser while working, you can access your document by starting Mozilla Firefox and choosing the link in the Favorites menu.

You will have to search the information required for the task in the internet. Please open other browser windows for your research.
You and the other participant have to coordinate who works on which part of the task. For this purpose, you can use the chat window that will be displayed in the upper left part of the screen (Screenshot 3). You and the other participant can communicate via the chat window during the whole 30 minutes. Please use only the chat window for your communication.

Please mark unambiguously, which participant has written which part of your document, as your individual evaluation and your bonus payment will depend on this. Please identify the respective parts in the document with “Participant 1” and “Participant 2”. Your participant number will be displayed in the first line on your chat window.

It is not allowed to reveal information in the chat or in the document that allows inferences on your identity. If you violate this rule, we have to exclude you from the experiment and all payoffs.
After you have finished the task, you have the opportunity to send a message to the evaluator. On the desktop of your computer, you find a WordPad document in which you can enter your message. Please open the message file after the working task has ended and you are asked to do so. You can save your message by clicking on the disk symbol (Screenshot 4).

If you have questions before and during the task, please raise your hand. We will come to your desk and help you.

Message to the Supervisor

Please type in you number here (Participant 1/2): ___
How do you assess your performance in the task relative to the other participant (please tick one option):
   _____ higher than the other participant
   _____ equal to the other participant
   _____ lower than the other participant

Please enter your message to the evaluator here: ______
Feedback: Bonus Payments

[Treatment TRANSPARENCY]
You receive a bonus payment of _____ % of the average bonus payment.
This accounts for _____ ECU.

[Treatment NON-TRANSPARENCY]
You receive a bonus payment of _____ ECU for the working task.

Instructions: Second Part
The second part of the experiment starts now.
In this part, you are matched with the same participant as in the first part of the experiment. In the second part, you and the other participant both receive an endowment of 100 ECU.
Each participant can decide whether he/she wants to contribute part of the endowment to a joint project. The sum of contributions will be multiplied by the factor 1.5 and equally distributed among both participants.
Your payoffs in the second part of the experiment are calculated as follows:

\[
100 \text{ ECU} - \text{Your contribution} + 1.5 \cdot \text{sum of all contributions, divided by 2} \\
\hline
= \text{Payoffs in the second part}
\]
You will be informed about your payoffs from the second part at the end of the experiment.

Instructions: Third Part
The third part of the experiment starts now. You receive an endowment of 40 ECU and can decide whether you want to transfer part of this endowment to your evaluator of the working task from the first part of the experiment. Any transfer will be doubled. The other participant that you were matched with before has to make the same decision.
That is, your payoffs in the third part of the experiment are calculated as follows:
Payoff = 40 ECU – transfer
The payoffs of the evaluator in the third part of the experiment are calculated as follows:
Payoff evaluator = 2 \cdot \text{sum of transfers of the two participants}

C.2 Instructions for Supervisors

Instructions: General Information
Welcome to the experiment! From now on, please do not communicate with other participants. If you have a question, please raise your hand! We will come to you and answer your question. If you violate these rules, we have to exclude you from the experiment and all payoffs.
In this experiment you can earn money. We use ECU (Experimental Currency Unit) as the laboratory currency. At the end of the experiment, your payoff in ECU is converted into Euro and paid out in cash. The exchange rate is 20 ECU = 1 Euro. The experiment consists of three parts. After the experiment, you receive the sum of payoffs from the parts that are relevant for you. In addition, you receive 18 Euros for your participation in the experiment, which is paid out at the end regardless of the decisions.

Instructions: First Part

In the first part of the experiment, you will be matched with two other participants. During the next 30 minutes, these two participants will work on the following task:
Which are the best and the worst homepages on electric cars you can find? Why? List the best argument for and against electric cars.
The two participants will jointly prepare a document, in which the results of their research are summarized. They will enter their texts online in a joint document on the “Google Docs” website. They have to list all references used (it is sufficient to paste the hyperlinks of the references in the document). The text has to be written in German.Copied parts have to be identified.
After the task has ended, the two participants have the possibility to send you an anonymous message with a self-assessment of their performances.
Your task is to rank the participants’ performance. For this purpose, you receive an excerpt of each participant’s contribution to the summary. The assignment of identical ranks is not possible. For the task, the two participants will receive an individual bonus payment depending on your evaluation. The average bonus is fixed. Your evaluations are anonymous. The identity of the other participants is confidential, and no other participant will be informed about your identity. During the 30 minutes the participants work on the task, you should make yourself familiar with the task on the internet.

Feedback: Supervisor

The higher performance in the working task was achieved by (please mark):

_____ Participant 1
_____ Participant 2

Instructions: Second Part

The second part of the experiment starts now.
In this part of the experiment, you will make no decisions and receive no payoffs.

Instructions: Third Part

The third part of the experiment starts now.
In this part of the experiment, each of the two participants you have evaluated in the first part receives an endowment of 40 ECU. Participants can now decide whether they want to transfer part of their endowments to you. Any transfer to you will be doubled by the experimenter. You will make no decision in the third part of the experiment.

Your payoffs in the third part of the experiment are calculated as follows:
Payoff = 2 \cdot \text{sum of transfers of the two participants.}

Payoffs of each participant in the third part of the experiment are calculated as follows:
Payoff participant = 40 \text{ECU} – \text{transfer}