



## Management Science

Publication details, including instructions for authors and subscription information:  
<http://pubsonline.informs.org>

### Bonus Payments and Reference Point Violations

Axel Ockenfels, Dirk Sliwka, Peter Werner

To cite this article:

Axel Ockenfels, Dirk Sliwka, Peter Werner (2015) Bonus Payments and Reference Point Violations. *Management Science* 61(7):1496-1513. <http://dx.doi.org/10.1287/mnsc.2014.1949>

Full terms and conditions of use: <http://pubsonline.informs.org/page/terms-and-conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact [permissions@informs.org](mailto:permissions@informs.org).

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2014, INFORMS

Please scroll down for article—it is on subsequent pages



INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

# Bonus Payments and Reference Point Violations

Axel Ockenfels

University of Cologne, D-50923 Köln, Germany, [ockenfels@uni-koeln.de](mailto:ockenfels@uni-koeln.de)

Dirk Sliwka

University of Cologne, D-50923 Köln, Germany; CESifo, 81679 Munich, Germany;  
and IZA, 53113 Bonn, Germany, [dirk.sliwka@uni-koeln.de](mailto:dirk.sliwka@uni-koeln.de)

Peter Werner

University of Cologne, D-50923 Köln, Germany, [peter.werner@uni-koeln.de](mailto:peter.werner@uni-koeln.de)

**W**e investigate how bonus payments affect the satisfaction and performance of managers in a large multinational company. We find that falling behind a natural reference standard for a fair bonus payment (a “reference point violation”) reduces satisfaction and subsequent performance. The effects are mitigated if information about one’s relative standing toward the reference point is withheld. A model and a laboratory experiment provide complementary insights and additional robustness checks.

**Keywords:** reference points; bonus payments; job satisfaction; job performance; transparency

**History:** Received April 10, 2013; accepted February 7, 2014, by Uri Gneezy, behavioral economics. Published online in *Articles in Advance* September 8, 2014.

## 1. Introduction

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 the satisfaction and the subsequent performance of the affected managers. As one consequence, supervisors tend to compress the distribution of bonuses at the reference point. Our hypothesis, based on straightforward theoretical reasoning, is that the effects can be substantially mitigated if a manager’s relative standing toward the reference standard is made less transparent. Making use of a cross-subsidiary difference in system handling with respect to how bonus payments are communicated to managers, as well as of laboratory data, we find strong supportive evidence for the critical role of reference point transparency.

Harmful effects of reference point violations in work settings, including falling behind a reference standard of “fair” wages, have been noted before. Regarding satisfaction, Bewley (1999, p. 43) concludes from his surveys of U.S. firms that “within a company, pay inequity offends (indeed, sometimes outrages) employees and destroys trust.” Regarding performance, Akerlof and Yellen (1990) introduce the “fair wage-effort hypothesis,” postulating that workers withdraw effort as their actual wage falls short of the reference standard for a fair wage. Yet, so far, only a few studies have investigated the impact of reference-dependent preferences and bonus 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.

We make use of a unique data set that enables us 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 salient and unique reference point for a fair bonus payment, largely consistent with behavioral economics approaches as discussed below. Also, our data allow us to match demographic characteristics, salaries, and bonus payments from personnel records with survey responses of managers in Germany (where the headquarters are located) as well as the United States. 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. However, the bonus payment is set by the supervisor as a percentage of the individual manager’s target (the “bonus percentage”) so that the actual payment typically differs from the manager’s target.<sup>1</sup> A manager’s bonus percentage is restricted by his

<sup>1</sup> The company has more than 100,000 employees worldwide. The largest subsidiaries are located in Germany and the United States.

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 should be chosen between 80% and 110%. Better-rated managers should be assigned 110% or more, and worse-rated managers should be assigned less than 80%. Because of the supervisors’ flexibility in assigning bonus percentages, managers who have the same performance rating may differ strongly in their individual bonus percentages. At the same time, 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. Hence, by increasing a manager’s bonus above 100%, another manager’s bonus must fall below 100%, and any deviation from 100% necessarily creates inequality among the managers in her department with respect to how they stand relative to their 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 standards proposed by two widely used types of reference-dependent preferences. The first type comprises *social* comparison models such as those in Fehr and Schmidt (1999) and Bolton and Ockenfels (2000). These models assume a particular notion of a reference standard, implying that people dislike falling behind others. Formally, many models entail a discontinuity in the slope of the utility function at the reference standard, which is the average group payment in Bolton and Ockenfels (2000) and the payment to another person in Fehr and Schmidt (1999). 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 their respective budgets (i.e., a bonus percentage of 100%). Yet if she 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, thus leading to socially unfavorable inequality measured in terms of bonus percentage. Hence, the 100% bonus percentage is a natural social comparison standard for bonus payments.

Based on the seminal work of Kahneman and 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. Bell (1985), Loomes and Sugden (1986), and Köszegi and Rabin (2006) further develop the essential intuition of Kahneman and Tversky’s work, postulating that people dislike falling behind their recently held expectations about outcomes. In the company we study, a manager who has learned that he is 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 or perceived entitlement for “fully meets expectations” managers.

In social contexts, Bolton (1991), Fehr and Schmidt (1999), and Bolton and Ockenfels (2000), 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 a 100% bonus percentage will strongly negatively affect a manager’s *satisfaction*, whereas assigning more yields a small effect, if it yields an effect at all. (We explicate the similarity between both concepts, and their implications for the assignment of bonus payments, in a formal model described in Online Appendix C. All online appendices are available for download in a single file via the following link: [http://ockenfels.uni-koeln.de/fileadmin/wiso\\_fak/stawi-ockenfels/pdf/publications/OSW\\_Online\\_Appendix.pdf](http://ockenfels.uni-koeln.de/fileadmin/wiso_fak/stawi-ockenfels/pdf/publications/OSW_Online_Appendix.pdf).)

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, p. 256) base their fair wage-effort hypothesis (that reference point violations reduce effort) on 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, Dufwenberg and Kirchsteiger 2004, Falk and Fischbacher 2006).<sup>2</sup> On the other hand, one might think that (the possibility of) falling below

Our study covers the incentive scheme for 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 those who are assigned bonuses as *managers* and to those who assign bonuses as *supervisors*. Managers in upper hierarchical levels can also be supervisors (who receive and assign bonuses), something we will exploit in our analyses in §4.

<sup>2</sup> Indeed, many laboratory and some field studies, starting with Fehr et al. (1993), suggest that reciprocity can play an important role in labor relationships. See Cooper and Kagel (2014) for a recent survey on other-regarding preferences, and Gächter and Fehr (2002), who survey laboratory labor market research. For example, Nagin et al. (2002) find in a field experiment that workers who reported

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 (Lazear and Rosen 1981).<sup>3</sup>

Summing up, we hypothesize that falling below the reference point strongly diminishes satisfaction. Because theory does not yield unambiguous predictions regarding the impact of reference point violations on performance, we refrain from formulating a hypothesis about performance. However, we expect that any satisfaction and performance effects as a result of reference-dependent preferences that we observe in Germany are mitigated in the United States. 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 United States. This is because labor regulations in Germany require a higher degree of transparency in compensation systems. Because bonus percentages cannot easily be inferred from absolute bonus payments and other data available to U.S. managers, we hypothesize—backed by a formal model—that the lack of transparency in the U.S. system diminishes the role of the 100% reference point.

Overall, our data strongly support the prominence of the 100% reference point. Transparent violations of the 100% reference point significantly reduce managers' satisfaction and their subsequent performance. We find no reference-dependent satisfaction and performance effects for the company's managers in the United States, who do not know their bonus percentage but otherwise operate under a basically identical bonus system. Finally, we observe that bonuses are significantly more compressed at the reference point in Germany than in the United States, which indicates that German supervisors partly anticipate the detrimental effects of reference point violations. A complementary lab

experiment supports our conclusions and yields further insights about potential underlying behavioral mechanisms.

Several studies investigate the impact of reference points on behavior in natural field environments.<sup>4</sup> Mas (2006) shows that police performance is sensitive to pay raises that are compared to the reference points set by final offer arbitrations. Kwon and Meyersson Milgrom (2009) analyze exogenous changes in workers' relative wages during mergers and acquisitions and find that status in the workplace affects turnover decisions. Chen et al. (2010) provide evidence that relative comparisons affect cooperative behavior within an Internet community. In a large-scale field experiment conducted with salespeople from a furniture company, Barankay (2012) observes that providing information about relative rankings affects performance negatively, although the rankings do not influence the employees' actual remuneration.

More closely related to our paper are studies conducted by Card et al. (2012), Gartenberg and Wulf (2013), and Cohn et al. (2014). In a controlled field experiment, Card et al. trigger social comparisons by informing employees of the University of California about the existence of a website that contains information about the wages of all employees. Among other things, the authors find that employees are most interested in the wages of their coworkers and that the job satisfaction of employees whose payments are below the median wage in their unit decreases significantly. Gartenberg and Wulf use a sample of Fortune 500 companies to test whether social comparisons affect wage setting for top executives. Indeed, the authors find a positive correlation of executive wages with peer compensation. Moreover, after a change in pay disclosure rules that made information sharing about peer wages easier within geographically dispersed units, the positive correlation becomes stronger in these units. Finally, Cohn et al. examine the role of wage comparisons in a field experiment. They decrease the fixed hourly wages in teams of two workers with identical tasks. Cutting the wages of both workers results in a significant performance decline. Yet if the wages of only one worker are reduced, the negative

a negative attitude toward their employer were more likely to behave opportunistically. Moreover, Neckermann et al. (2014) observe that assigning awards for voluntary work had a positive impact on the subsequent performance of call-center agents in tasks that were unrelated to the award. For more recent surveys of the literature on reciprocity at the workplace, see Fehr et al. (2009) and Rebitzer and Taylor (2011). 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 §5).

<sup>3</sup> Demougin and Fluét (2003) and Grund and Sliwka (2005) show in theoretical models that the incentive effect of tournaments may become stronger when individuals are inequity averse.

<sup>4</sup> Earlier studies on the effects of relative wage comparisons include Pfeffer and Langton (1993), who use survey data from U.S. universities and find that job satisfaction, productivity, and the tendency to work collaboratively are negatively related to measures for wage dispersion, thus providing indirect evidence that workers' orientation on reference points may have detrimental effects. Moreover, Levine (1993) uses survey data from workers in the United States and Japan and concludes that relative wages are positively correlated with pay satisfaction and self-reported commitment. In a sample from a British household panel survey, Clark and Oswald (1996) find that, controlling for absolute payments, job satisfaction declines with higher reference incomes. Clark et al. (2008) review the growing body of research about the importance of reference incomes on survey measures of life and job satisfaction.



response of that particular worker is much stronger than in the case of symmetric wage cuts, highlighting the impact of social comparisons within teams.

Our paper makes several contributions to the literature. For one, our study provides large-scale evidence for the impact of a naturally occurring reference point on the effectiveness of incentive systems in a multinational company. Specifically, although most related studies have investigated the role of relative comparisons with respect to fixed wages or total compensation, we study the potential detrimental effects of performance-related pay aimed at creating work incentives. Moreover, we focus on the interaction of a fixed total bonus budget and the degree of bonus transparency, which are important design choices in many relevant business environments that employ bonus systems. By providing complementary field evidence, laboratory evidence, and new theory, we illustrate the important and robust role of reference point transparency for satisfaction, performance, and evaluation. Given the high relevance and the monetary significance of bonus plans in practice, proper system design is of critical importance for both firm performance and the well-being of employees.

Section 2 describes our data and shows that the distribution of bonus percentages in Germany is much more compressed at 100% in Germany than in the United States. Section 3 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. In the model we assume that agents have a utility function with a kink at some reference standard. This reference standard can either be a social comparison standard or a manager's prior expectation about the bonus. 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 4 investigates the impact of reference point violations on performance. In §5, we provide an additional robustness check of our main results with the help of a controlled laboratory environment that captures relevant features of the bonus system in the field. The experiment also complements the indication in the field data that the performance effect is partly due to negative reciprocity. Section 6 discusses the findings and concludes.

## 2. 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 United States (summer 2008). Because the survey was part of a larger study, managers

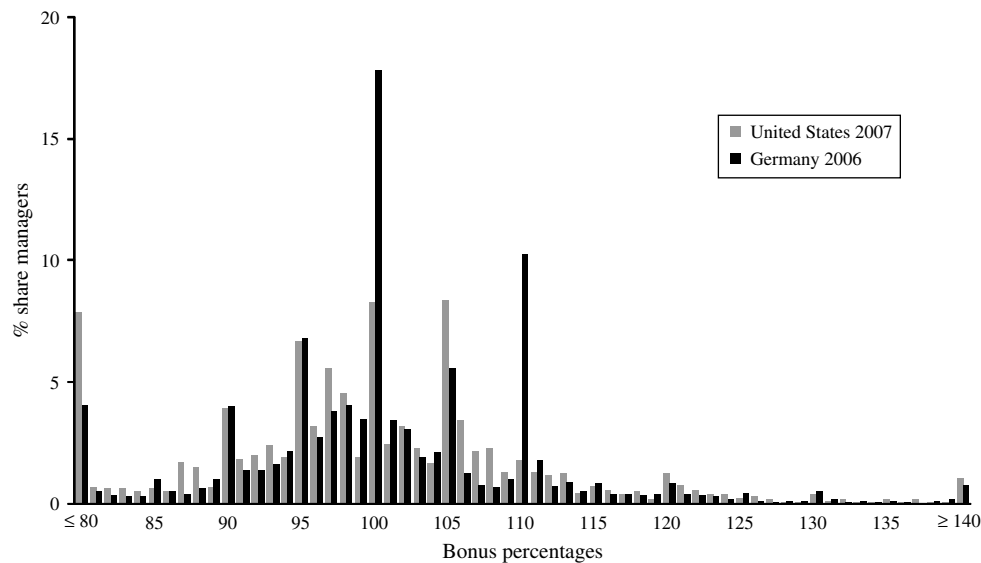
were asked some 60 questions about workplace characteristics and other aspects of work, including job satisfaction. We independently collected data about performance evaluations, compensation, demographic characteristics, and department affiliations of the managers over the years 2004–2006 (Germany) and 2004–2007 (United States) from the personnel records of the firm. The technical environment allowed us to connect background data with the survey answers of the individual managers in a way that guaranteed the anonymity of the participants of the survey (see Online Appendix A). All together, 4,997 managers took part in the survey (3,122 in Germany and 1,875 in the United States), which corresponds to a participation rate of 59% and 41%, respectively.<sup>5</sup>

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 United States. 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 United States have common organizational structures, and the bonus systems in Germany and the United States operate under basically identical formal rules. There are also no systematic differences in informal rules, e.g., with respect to goal setting, or in the communication content, timing, and procedures, with one important exception: Because of 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 United States, on the other hand, managers are only told their bonus payment.<sup>6</sup>

Salaries and absolute bonus payments of German and U.S. managers are roughly comparable during

<sup>5</sup> The lower response rate in the United States can be attributed to the fact that the U.S. survey was run at a time when a larger fraction of the employees was on holiday. Comparing participating and nonparticipating managers, most demographic and compensation data are very similar (see Table B1 in the online appendix). However, the share of women and old managers among German survey participants is larger than among German nonparticipants (see Table B2 in the online appendix for more details). We will explain in §3 that this does not affect our conclusions.

<sup>6</sup> 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 that of the relevant division, subject to exceptions and discretion. Also, 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 U.S. manager to learn his individual budget. In fact, the open survey answers of U.S. 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 [percentage] was applied."

**Figure 1** Distribution of Bonus Percentages in Germany and the United States

the period of analysis.<sup>7</sup> In 2006 (2007), the average fixed salary of a manager in Germany (in the United States) was USD 107,700 (USD 113,200). The bonus payments are substantial, as they account for some 20% of fixed salaries on average (USD 22,200 in Germany and USD 22,700 in the United States).

There are first indications of the importance of the 100% reference point when we look at how bonuses in Germany and the United States are distributed. Figure 1 shows the bonus percentage distributions in Germany for 2006 and in the United States for 2007, the respective years of our survey (the distributions are very similar in the other years). Compared with the U.S. data, bonus percentages in Germany are much more compressed toward the 100% reference point. The difference in the distributions is statistically significant (two-sided Kolmogorov–Smirnov test,  $p < 0.001$ ). Most strikingly, in Germany, there is a significant peak at 100%, with 17.6% of the managers receiving exactly their target bonus, whereas in the United States only 8.3% of the managers receive a 100% bonus percentage ( $p < 0.001$ , two-sided  $\chi^2$  test). In addition to the 100% reference point, 80% and 110% have special prominence in the bonus system because they constitute the upper (lower) threshold for bonus payments of managers with a “below average” (“above average”) rating. Correspondingly, the next largest peak in Germany is at 110%, which is the feasible bonus closest to 100% for the “above average” managers. In the United States, on the

other hand, 110% exhibits no specific prominence—only 4.3% of all “above average” managers in the United States receive exactly 110%, which is approximately 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 2.2% (11.6%), compared with 6.6% (13.4%) in the United States, (weakly) significantly smaller at  $p < 0.001$  ( $p = 0.054$ ). Looking at the whole distribution, the standard deviation of bonus percentages observed in Germany is 12.2 and is therefore significantly smaller than the standard deviation of 15.1 observed in the United States ( $p < 0.001$ , two-sided Levene test for differences in variances). We summarize our findings in the following observation.

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

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 United States. Because paying managers less than 100% significantly reduces satisfaction and performance if the reference point violations are transparent (see §§3 and 4), there is more reason for compression toward the reference point for “fully meets expectations” managers in Germany than in the United States (as predicted by our model

<sup>7</sup> Although U.S. managers receive somewhat larger base salaries, the share of managers with bonuses less than USD 10,000 is higher in the United States 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 U.S. managers on the lowest hierarchy level.

**Table 1** Determinants of Job Satisfaction (Ordered Probit Models with the “Job Satisfaction” Score as the Dependent Variable)

	Germany					United States				
	D1	D2	D3	D4	D5	US1	US2	US3	US4	US5
Salary (000's)	0.000 [0.004]	0.000 [0.004]	0.001 [0.004]	0.001 [0.004]	0.002 [0.004]	−0.001 [0.003]	−0.001 [0.003]	−0.001 [0.003]	−0.001 [0.003]	−0.005 [0.004]
Bonus payment (000's)	0.001 [0.009]	0.000 [0.009]	−0.002 [0.009]	−0.002 [0.009]	−0.006 [0.011]	0.001 [0.005]	0.002 [0.004]	0.002 [0.005]	0.002 [0.005]	0.011* [0.006]
Bonus percentage	0.016*** [0.003]	0.008** [0.004]				0.012*** [0.004]	−0.002 [0.006]			
Above meets		0.143* [0.084]		0.270*** [0.091]			0.233** [0.107]		0.232** [0.107]	
Below meets		−0.791*** [0.194]		−0.419* [0.219]			−1.123*** [0.257]		−1.150*** [0.329]	
Positive deviation from bonus budget (=100%)			0.005 [0.004]	−0.005 [0.005]	0.004 [0.011]			0.005 [0.006]	−0.001 [0.008]	−0.015 [0.028]
Negative deviation from bonus budget (=100%)			−0.032*** [0.004]	−0.025*** [0.006]	−0.027*** [0.007]			−0.019*** [0.005]	0.003 [0.009]	0.004 [0.011]
Sample	All	All	All	All	“Fully meets”	All	All	All	All	“Fully meets”
Observations	2,044	2,044	2,044	2,044	1,506	956	956	956	956	599
Pseudo- $R^2$	0.035	0.038	0.038	0.040	0.029	0.031	0.034	0.031	0.034	0.034
Log-likelihood	−3,072	−3,063	−3,063	−3,057	−2,275	−1,455	−1,450	−1,454	−1,450	−927

Notes. Robust standard errors clustered on the level of supervisors are given in brackets. The dummy variables *above meets* (*below meets*) refer to higher (lower) performance ratings than “fully meets expectations.” 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.

\*Significant at the 10% level; \*\*significant at the 5% level; \*\*\*significant at 1% level.

described in §3).<sup>8</sup> And the more managers who receive more than 100%, the larger the budgetary pressure to violate the other managers’ reference point, which is harmful when the violation is transparent. In fact, we observe that the performance rating “above average” is assigned significantly less often in Germany (see Table B3 in the online 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 the fair wage-effort hypothesis of Akerlof and Yellen (1990), as well as with the work of Frank (1984) 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).

### 3. How Does the Bonus Matter for Job Satisfaction?

In this section, we investigate whether there is an asymmetric response in the managers’ well-being to positive

and negative deviations from the 100% reference point as postulated by the models of reference-dependent preferences outlined in §1. A commonly used proxy for job satisfaction is the degree of approval to the statement: “I am very satisfied with my job.”<sup>9</sup> In our survey, managers could respond to this statement 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 responses in the United States and Germany are rather similar, with a mean of 5.29 (SD = 1.25) in Germany and 5.24 (SD = 1.25) in the United States (see Table B4 in Online Appendix B). We estimate ordered probit models with job satisfaction as the dependent variable for the years in which the respective survey studies took place, 2006 (Germany) and 2007 (United States). We include controls for age intervals, gender, firm tenure, promotion in the previous business year, and the company unit. Moreover, in all specifications, we report robust standard errors clustered on the level of supervisors to account for interdependence within teams. In some specifications we also control for the effects of performance ratings. Table 1 reports the results.<sup>10</sup>

<sup>9</sup> See, for instance, Freeman (1978), Clark and Oswald (1996), and Hamermesh (1977, 2001). Frey and Stutzer (2002) and Luttmer (2005) discuss the use of self-reported measures of well-being in economics.

<sup>10</sup> The results are similar when we use ordinary least squares (OLS) regressions on a unit normal transformation (e.g., see Freeman 1978) of the satisfaction scores.

<sup>8</sup> In line with this notion, U.S. supervisors tend to deviate more from a manager’s individual budget: in the years relevant to our survey, the average share of “fully meets expectations” managers per supervisor who receive bonuses lower than their budgets is 81.9% in the United States compared with 61.0% in Germany.



In our first specification, we focus on the impact of absolute bonus payments and bonus percentages. Because a manager's bonus budget depends on his position in the company, his salary, and the performance of both the company as a whole and the manager's relevant division, bonus payments and bonus percentages are not collinear,<sup>11</sup> 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, whereas the coefficient of bonus percentages is highly significant and positive for both Germany (Model D1) and the United States (Model US1).<sup>12</sup> Hence, despite their substantial economic relevance, absolute bonus payments do not correlate with job satisfaction, although bonus percentages do.

When we include ratings dummies (Model 2), there is a strong connection between performance ratings and job satisfaction in Germany as well as in the United States: managers with a rating better than "fully meets expectations" are significantly happier, and those with a rating worse than this grade are significantly less happy. More importantly, the positive effect of bonus percentages remains significant for Germany, yet the U.S. coefficient of the bonus percentage becomes insignificant. This observation is in line with the difference in transparency in Germany and the United States: because managers in the United States 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, we do not find a significant correlation between one's relative position toward the reference point and satisfaction. This suggests that the effect of percentages observed in Model US1 is driven mainly 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.

As we have pointed out in §1, 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 affect

or only marginally affects 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:

$$\text{JobSatisfaction}_i = \alpha + \beta \cdot X_i + \gamma \cdot (z_i - 100) \cdot I_{\{z_i > 100\}} + \delta \cdot (100 - z_i) \cdot I_{\{z_i < 100\}} + \varepsilon_i,$$

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 bonuses, we expect that  $\delta$  is significantly negative and  $\gamma$  is positive but small.

Consistent with the hypothesis, we find that falling behind the 100% bonus comes along with significantly smaller satisfaction levels, whereas increases above 100% do not increase self-reported satisfaction in Germany (see Model D3). The asymmetric effect of being below and above the reference point is also apparent in the United States (Model US3), even though U.S. managers do not know their bonus percentages. Again, the likely reason is that bonus percentages are correlated with performance ratings, so U.S. 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 coefficient for deviations below 100% becomes insignificant in the U.S. sample while remaining significant in Germany.<sup>13</sup> Moreover, as a rough orientation for the economic size of the estimated effects in the German sample, we calculate the marginal effects of our deviation variables from Model D4. The estimates suggest that reducing the bonus percentages by 10 percentage points below the budget *ceteris paribus* decreases the probability to report a high job satisfaction score by 4.02 percentage points (satisfaction score = 7) and 5.75 percentage points (satisfaction score = 6), respectively.<sup>14</sup>

<sup>13</sup> 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 B5 in the online 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.

<sup>14</sup> We note that we interpret these effects only as a rough indication, because they are calculated under the assumption that all other independent variables (some of which are dichotomous) take their mean values and that bonus percentages are exogenously assigned.

<sup>11</sup> The correlation coefficient is  $r = 0.204$  in the German and  $r = 0.153$  in the U.S. sample.

<sup>12</sup> Absolute bonus payments and bonus percentages have both a positive and mostly significant impact when only one variable is included in the model. Also, the models 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 (e.g., see Clark 1999 and the references cited therein).



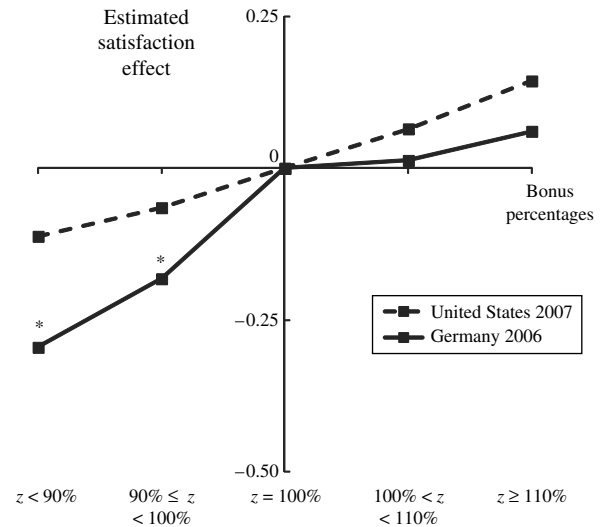
We also estimate Model 3 with a sample including only managers who receive the “fully meets expectations” rating. This yields similar results for the German and the U.S. sample as in the previous model (see Models D5 and US5).

As a further robustness check, we combine the two country samples and estimate a model with the same independent variables as Models D4 and US4 (Models D5 and US5). We additionally include the dummy *GER*, which is equal to 1 if the observation is from the German sample, as well as its interaction terms with the deviation variables, to measure whether deviations from the 100% budget have different effects in the two countries. The models replicate all our previous findings with respect to the effect of base salary, absolute bonus payments, and performance ratings.<sup>15</sup> Moreover, with respect to the negative deviations from the reference point, we observe in both models that the coefficient for the U.S. managers is insignificant, whereas the interaction effect that measures the response of German managers is negative and significant. This supports our conclusion that the negative effect of reductions below the 100% bonus is found only for German managers.

Following our discussion in §1, the statistical analyses above assumed that the reference point (i.e., the kink in the estimated function) is at 100%. 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 United States, including dummies for percentage intervals instead of bonus percentages and controls for performance ratings as well as supervisor fixed effects (see Table B7 in Online Appendix B). Figure 2 displays the estimated marginal effects of positive and negative deviations from the 100% reference bonus in Germany and the United States, controlling for the effects of absolute remuneration, demographic background, and the organizational unit of the manager. The reference group is the group of managers with a bonus of exactly 100% in both countries. Hence, by construction, the graph normalizes satisfaction at the level of managers with a 100% bonus in each country and displays the satisfaction effect of managers with other bonus percentages relative to this benchmark.

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 United States, too, both dummies below (above) 100% have negative (positive)

**Figure 2** Relationship Between Bonus Percentage and Job Satisfaction Relative to Managers with a 100% Bonus Percentage



**Notes.** 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 B7 in the online appendix). Significant interval dummies (at the 5% level) are indicated by an asterisk. The reference group consists of managers in each country who receive exactly a 100% bonus.

signs, but here, all interval dummies are insignificant.<sup>16</sup> For Germany, estimated interval dummies exhibit the expected kinked shape at the reference point.<sup>17</sup> No such pattern is observed in the U.S. data.

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%. (Figure B1 in the online appendix plots the results.) It turns out that the goodness of fit measured by the  $R^2$  (within) value is maximized with a kink at

<sup>16</sup> 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 6). Hence, last year’s bonuses cannot easily serve as a reference point for expectations regarding subsequent bonuses. Also to capture the relative importance of social comparisons and personal expectations, we ran an additional model in which we use the same variables as in Model D5 and include variables for *changes in the absolute bonuses* from 2005 to 2006 as well as for *changes in bonus percentages*. We include only managers who received a “fully meets expectations” rating in both years to avoid confounding effects from changes in the performance rating and the resulting shifts in bonus payments and percentages. Yet all variables capturing historical bonus payments turn out to be insignificant while the significant negative effect of deviations below 100% remains robust.

<sup>17</sup> The described relation remains robust if the analysis is restricted to managers rated “fully meets expectations.”

<sup>15</sup> The results are displayed in Table B6 in the online appendix.

100% or 101%, which highlights the special importance of the reference point.

As a final check, to rule out the possibility that our results in the German sample are influenced by the demographics of the survey participants (see Footnote 5), we estimate Model D4 separately for male and female managers and for managers who are 44 years or younger (the median age in our sample) or are older than 44 years. Tables B8 and B9 in the online appendix list the results. Again, we find the previously reported asymmetric effect of positive and negative deviations from the 100% budget in all of these subsamples, suggesting that reference point violations affect all groups of managers.<sup>18</sup>

That said, we note that 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 asymmetry of positive and negative deviations from a manager's budget 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 U.S. 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 robust significant relation between percentages and satisfaction in the United States, 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 experiments (as discussed in §5) and field experiments (see our discussion of the related literature in §1).

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 creates unhappiness as a result of anticipated utility effects. We therefore test whether and how deviations from the 100% bonus are related to the future career progression of a manager. To do this we make use of the remuneration data of the German managers for the years 2004–2006 and 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 1 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 association between below-average bonuses and promotions. However, we do not find a larger effect of negative deviations than of positive deviations in any specification. In fact, the size of the coefficient for positive deviations always exceeds that for negative deviations: for example, in Model 1, a manager with a “fully meets expectations” rating and a 110% bonus has a 8.0% higher probability of being promoted than managers with a 100% bonus, whereas the probability of promotion for a manager with a 90% bonus decreases by only 5.0% compared with the reference group.<sup>19</sup> Furthermore, we neither find a quantitatively stronger effect of negative deviations from 100% bonus 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 a salary increase of EUR 75 less for the year 2006. However, every percentage point above 100% predicts an estimated salary increase of EUR 131 (see Table B10 in Online Appendix B). We conclude that the asymmetric pattern of satisfaction and dissatisfaction with the bonus does not follow the corresponding signaling values of bonuses. Hence, career signaling cannot explain the estimated shape of job satisfaction.

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.

<sup>18</sup> If we compare the absolute sizes of the coefficients in our models for job satisfaction for positive and negative deviations using two-sided Wald tests, we can reject the null hypothesis that  $\gamma = -\delta$  in almost all cases for German managers: in Models D3 and D4 presented here, in Model 1 in Table B5, in Models 1 and 2 in Table B8, and in the model in Table B12, the significance level is 1%; in Model 2 in Table B5, it is 2%; and in Models 1 and 2 of Table B9, it is 4% (the tables can be found in Online Appendix B). The only model where the difference is not significant at 5% is Model D5, which is driven by the high standard error of the insignificant positive coefficient in this specification.

<sup>19</sup> The results are reported in Table B10 in Online Appendix B. 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.

To further show that reference-dependent preferences organize Observations 1 and 2, we develop a formal model that combines a standard approach to reference-dependent social utility with a framework to analyze subjective performance evaluations. In particular, managers are assumed to have a utility function with a kink at a reference standard, which can equally be interpreted as a social comparison standard and as an aspiration level for the bonus. When the reference standard is the average bonus paid to colleagues, this boils down to a Fehr and Schmidt (1999) type of utility function, where the comparison standard is the average income, such as in Bolton and Ockenfels (2000). When the reference standard is the agent's aspiration level, it can be interpreted as a model of loss aversion with respect to the manager's individually perceived entitlement toward a 100% bonus. Supervisors are assumed to prefer accurate evaluations and to also care to some extent about their managers' well-being (such as in Prendergast and Topel 1996, Prendergast 2002). 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 directly implied by these standard approaches to model reference-dependency and subjective performance evaluations. The model is described in detail in Online Appendix C; 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 utility is three-fold: First, the bonus may affect material well-being. Second, the bonus payment reveals information about one's own standing relative to the reference standard. And third, the manager evaluates his own bonus payment relative to his conditional expectation about the reference bonus. If the prior is perfectly precise (i.e., the manager knows exactly his relative position), his utility exhibits a kink at the reference bonus. However, we show that 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. In other words, we show that even managers who in fact care about a reference standard behave as if they do not care, as the agents' information about their relative standings becomes vague. This mirrors 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 and 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 the relative standing are nontransparent 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 Observation 1.

#### 4. How Do Reference Point Violations Affect Job Performance?

Although reference point violations negatively affect employee perceptions, they might be useful for generating 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: 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 reduced department performance and thus also negatively affect the department supervisor's performance rating in the subsequent year. The supervisors' rating comes from a performance appraisal by their own supervisors and depends neither on the evaluation by peers nor subordinates. As in the linear specifications of our job satisfaction models (see Table B7 in the online appendix), we use a unit normal transformation of the performance rating to minimize distortions because the rating is measured on an ordinary scale.

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 work group effects. We estimate the following specification:

$$Rating_{st+1} = \alpha + \beta \cdot DevRef_{st} + \gamma \cdot X_{st} + \mu \cdot Y_{st+1} + a_s + \varepsilon_{st}$$



**Table 2** Performance Effects of Reference Point Violations (Dependent Variable: *Adjusted Supervisor Performance Rating in Subsequent Year*)

	Germany		United States	
	D1	D2	US1	US2
Share of managers with bonus percentage < 100% among “fully meets” managers	−0.572*** [0.217]	−0.676*** [0.211]	0.009 [0.178]	−0.087 [0.165]
Controls for rating distribution	No	Yes	No	Yes
Observations	468	468	504	504
R <sup>2</sup> (within)	0.05	0.09	0.05	0.08

Notes. All models include supervisor fixed effects. Robust standard errors clustered on the level of the supervisors are given in brackets. 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.”

\*\*\*Significant at the 1% level.

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

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 pushed below the 100% bonus percentage in a given year, the worse 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  $t$ . For one, although 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 on 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 disproportionately large share of the total budget be spent on the excellent managers.<sup>20</sup> 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.<sup>21</sup>

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 an additional 24.6% of all other managers when changing her evaluation behavior in this way.<sup>22</sup>

That such reference point violations hurt a supervisor's rating raises the question of why reference points are violated in the first place (although this happens much less than when violations are nontransparent). Our interviews with the company's human

<sup>21</sup> 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.

<sup>22</sup> 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 33rd (66th) 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 the larger bonuses of high performers.

<sup>20</sup> If we estimate a simple OLS model (not reported here) with our measure for reference point violations as a dependent variable and include the number of managers per supervisor and the shares of the respective performance ratings as explanatory variables, the effects of the rating behavior are as expected: performance ratings higher (lower) than “fully meets expectations” increase (decrease) the budgetary pressure of the supervisor and thus affect the degree of reference point violation in the same direction. However, team size per se does not have an effect on reference point violation.



resources (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, however, may be out-of-equilibrium behavior: not every supervisor seems to realize the detrimental effect of reference point violations (although some apparently do). Consistent with this view, our research results led the company to redesign the rules by which bonuses are distributed, suggesting that neither all supervisor behavior nor the institutional design that we study were in equilibrium.

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.”<sup>23</sup>

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

The regression results are robust in alternative specifications for the German supervisors (see Table B11 in Online Appendix B). 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 (Model D5), the respective coefficient stays negative and highly significant for Germany. Also, 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 (Model D6) and for the

number of subordinates of a supervisor (Model D7).<sup>24</sup> 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 (Model D8). The coefficient for the frequency of reference point violations remains significantly negative in all specifications for Germany. Moreover, there is no significant effect of reference point violations in any of these specifications for the United States (not reported here).<sup>25</sup>

We summarize the main result of this section in the following observation.

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

In search of individual motives underlying the detrimental effect of reference point violations, we find some evidence that, in line with the fair wage-effort hypothesis, reciprocal reactions and attribution of intentions after reference point violations play a role. Specifically, we can make use of a survey item, “My supervisor evaluates my general performance appropriately,” which had to be rated on a seven-point scale from 1 (totally disagree) to 7 (fully agree). We hypothesized that the more a given reference point violation is perceived as “appropriate,” the less it can be interpreted as unfair or unkind, and thus the less it should be prone to a reciprocal response. This suggests that reference point violations are less harmful if the supervisor is considered a rather accurate evaluator.

The data confirm the prediction. In Table 3 we report the performance regression D2 (from Table 2), now interacting the share of reference point violations either with a dummy variable indicating that a supervisor belongs to the top quartile in terms of the perceived appropriateness (Model D3), or with the supervisor’s percentile in the ranking according to this criterion (Model D4). Both models show that the performance

<sup>23</sup> Other examples are as follows: “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 [percentage]-point below that is a disappointment”; and “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.

<sup>24</sup> 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.

<sup>25</sup> In addition, although we have only relatively few observations for this analysis, we can provide some indication that the results are unaffected by controlling for the extent to which team composition changes. For instance, one could imagine that bringing in new managers leads to more bonus payments below 100% (because new managers are natural candidates for lower bonuses) and as well as reduced performance because of the initial lack of human capital. We determine the percentage share of new managers on a team by comparing the teams of each supervisor in 2004 to the teams in 2005 and estimate an OLS model on the cross section of German managers for the year 2005 with the performance rating in 2006 as the dependent variable, additionally controlling for the share of newly assigned managers (model not reported here). Yet the coefficient of this variable is insignificant whereas the effect of our measure for reference point violation remains robust.

**Table 3** Performance Effects and Supervisor Accuracy (Dependent Variable: *Adjusted Supervisor Performance Rating in Subsequent Year*)

	Germany	
	D3	D4
Share of managers with bonus percentage < 100% among “fully meets” managers	−1.116*** [0.265]	−1.400*** [0.424]
Accurate supervisor × Share of managers with bonus percentage < 100%	1.117*** [0.303]	
(Percentile accuracy/100) × Share of managers with bonus percentage < 100%		1.119** [0.529]
Controls for rating distribution	Yes	Yes
Observations	439	439
R <sup>2</sup> (within)	0.16	0.12

Notes. All models include supervisor fixed effects. Robust standard errors clustered on the level of the supervisors are given in brackets. Control variables include the age and hierarchy level of the supervisor. A supervisor's rating distribution is measured by the share of the performance ratings “excellent,” “above average,” and “below average.”

\*\*Significant at the 5% level; \*\*\*significant at the 1% level.

effect is smaller when a supervisor's evaluations are perceived to be more appropriate. Moreover, the coefficients in Model D3 indicate that the detrimental effect of reference point violations vanishes entirely for those supervisors who provide the most appropriate evaluations.<sup>26</sup> Taking this information together, we conclude that because reference point violations trigger a negative performance effect only to the extent that supervisors are perceived to evaluate inappropriately, our results are consistent with the notion that negative reciprocity explains the performance effect.<sup>27</sup>

Our results consistently show a significant impact of transparent reference point violations on job satisfaction and performance. However, regarding the strong role of transparency in our data, one might argue that there are other, cultural differences between Germany and the United States that might contribute to our

<sup>26</sup> This effect is not driven by a lower general propensity of accurate supervisors to violate the reference points of their managers. In fact, supervisors whose adequacy is rated with a score of less than 6 assign on average 58.3% of their “fully meets expectations” managers less than 100%, whereas the corresponding value for supervisors with a score of 6 and higher is 53.6%. We also note that although the distribution of adequacy scores is comparable (with an average of 5.13 (4.96) and a standard deviation of 1.48 (1.65) in Germany (the United States)), neither variable becomes significant, and coefficients are small in similar models for the U.S. sample (not reported here).

<sup>27</sup> One might speculate, assuming that satisfaction depends on perceived accuracy and given that accuracy depends asymmetrically on deviations from 100%, that the satisfaction effect of reference point violations described in §3 disappears when we control for the perceived accuracy of the supervisor. This is not the case, however; when we include a manager's individual assessment of her supervisor in Model D4 from Table 1 (see Table B12 in Online Appendix B), it is positively and significantly correlated with the satisfaction score. Yet the significant kink at the 100% reference point remains.

observations. Although we cannot fully exclude this possibility, we emphasize that we have little reason to suppose that differences in culture play a crucial role in explaining the data. In line with our interpretation that transparency is the main driver, behavior and outcomes are roughly the same in both countries when we do not expect them to be affected by the “transparency gap” and differ when widely accepted behavioral models of reference-dependent behavior and performance evaluation would predict differences. For instance, as we have seen in §3, managers in the United States, too, do not like falling behind with respect to performance ratings (Model US3); yet, if we control for the information about bonus percentages contained in the ratings, no evidence for an asymmetric effect of deviations from the 100% bonus is left (Models US4 and US5). This indicates that the difference stems from the (non)transparency of bonus percentages and not from the absence of reference-dependent preferences in the United States. Another example is that managers' job satisfaction is similarly increasing after receiving a better rating than “fully meets expectations” (see Table 1).

Also, previous field and laboratory research suggests that cross-country effects between the United States and Germany are limited. For instance, studies on life satisfaction have indicated the importance of relative income comparisons both in the United States and in Germany (e.g., see Ferrer-i-Carbonell 2005, Luttmer 2005). Moreover, there is some evidence for measurement equivalence of job satisfaction measures in countries sharing similar cultural backgrounds such as Germany and the United States. For instance, Liu et al. (2004) find satisfactory measurement equivalence of a particular job satisfaction measure between western Europe and the English-speaking countries. Laboratory experiments typically find little difference in reference-dependent behavior between northern Europe and the United States, also suggesting that reference-dependent preferences are similar (e.g., see Brandts et al. 2004 for a cross-cultural study of cooperation in dilemma games and Oosterbeek et al. 2004 for a metastudy of bargaining experiments). In the next section, we use laboratory experiments to provide complementary evidence for the important role of reference point transparency.

## 5. A Laboratory Experiment

Based on a laboratory study, this section provides another robustness check for our observations and sheds further light on the underlying behavioral mechanisms. 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. Because the aim of

the experiment is to isolate behavioral mechanisms through which transparency has a negative effect on satisfaction and performance among managers in our field data, we abstract away from factors such as repeated game effects and cultural differences that may have an additional impact on evaluation behavior on the supervisor's side.<sup>28</sup> 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://docs.google.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 Online Appendix D for instructions). How the task was approached and how 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.<sup>29</sup> Supervisors and workers interacted anonymously.

We implemented a  $2 \times 2$  design to systematically analyze the interaction of bonus inequality (EQUAL and UNEQUAL) and transparency (TRANSPARENCY and NONTRANSPARENCY) 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% (EUR 24) of the average bonus budget, and the low-ranked worker received 80% (EUR 16). In the EQUAL treatments, both the high- and the low-ranked worker received 100% of the average budget (EUR 20). 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 NONTRANSPARENCY 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.<sup>30</sup> Next, each of

the workers received an extra endowment of EUR 2 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 for Economic Research with altogether 120 subjects, yielding 10 independent observations for each treatment. Participants were recruited by using the online recruitment system ORSEE (Online Recruitment System for Economic Experiments; Greiner 2004). The chat and the second and third parts of the experiment were implemented by using z-Tree (Fischbacher 2007). Every session lasted about two hours; the average payoff (SD) was EUR 28.02 (EUR 6.50) including a show-up fee of EUR 4 for workers and EUR 18 for supervisors. After each session, participants filled out a postexperimental questionnaire in which, among other things, 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 with 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 NONTRANSPARENCY treatment (all tests yield  $p$ -values  $> 0.1$ ). In the transparent condition, however, the relevant differences are statistically significant.<sup>31</sup> Moreover, negative deviations from the 100% bonus have a substantially stronger impact on satisfaction than positive deviations: workers with an 80% bonus report a satisfaction score that is 2.4 points lower than subjects with 100%, whereas receiving 120% increases satisfaction by only 0.9 points.

The second and third parts of the experiment investigate how the bonus from the first part of the experiment affects the workers' willingness to cooperate with each other and potential reciprocal reactions toward the supervisor. In particular, these parts may help to investigate whether potential negative responses of workers relate to explanations based on outcome-based fairness models and intention-based reciprocity. More

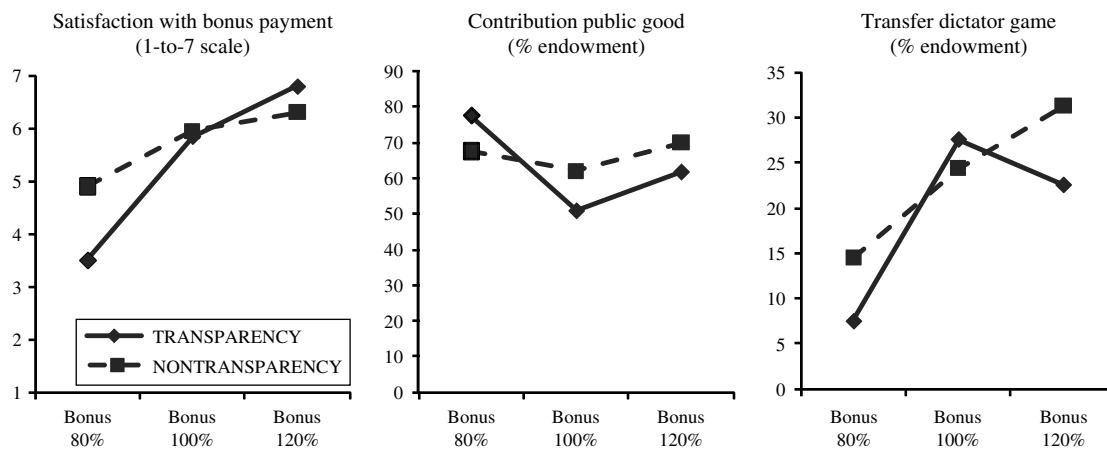
<sup>28</sup> See Harrison and List (2004) for the value in combining laboratory and field data; see, e.g., Kagel and Roth (2000) and Ariely et al. (2005) for a similar research strategy in economic engineering projects.

<sup>29</sup> The workers were also asked to send the supervisor a self-evaluation, which, however, turned out to have no effect on any of the variables we are investigating here. Moreover, all feedback we received from our laboratory supervisors suggested that they took the task of rating the agents' excerpts very seriously. Therefore, it seems unlikely that our results are much distorted by arbitrary assignments of ratings to the experimental workers.

<sup>30</sup> Both workers received an endowment of EUR 5, 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.

<sup>31</sup> Comparing subjects with 80% and 100% bonuses and subjects with 100% and 120% bonuses by using two-sided Mann-Whitney  $U$  tests yields significance levels of  $p = 0.002$  and  $p = 0.003$ , and comparing subjects with 80% and 120% bonuses by using a two-sided sign test yields a significance level of  $p = 0.002$ .



**Figure 3** Experimental Results (Averages per Treatment)

specifically, if workers are motivated to *reduce payoff inequality* among themselves after a transparent reference point violation, we would expect to see that workers paid below average 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.<sup>32</sup>

However, if workers use relative pay information to assess the fairness of their wage, which forms the basis for their reciprocal behavior toward the supervisor, and are motivated to *reciprocally punish* the supervisor if a given wage is considered as unjust, 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. We note that, for example, in the reciprocity model of Falk and Fischbacher (2006), 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 reducing satisfaction but also triggering reciprocal reactions.

The data indeed are in line with 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 NONTRANSPARENCY treatment. Workers with an 80% bonus transfer 7.5% of the endowment versus 27.3% of the workers with a 100% bonus ( $p = 0.011$ , 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).

<sup>32</sup> Interestingly, subjects with an 80% bonus in the transparent condition tend to contribute even 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 are attempting to catch up with the workers' "overall average effort" after their respective contributions to the team production in the first part of the experiment had been evaluated below average. All other comparisons are insignificant on conventional levels.

Finally, as in the case of pay satisfaction, we do not observe a statistically significant difference between any worker groups in the NONTRANSPARENCY treatment.<sup>33</sup> That is, only transparent reference point violations induce a significant negative reciprocal response toward 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 in theory and field and laboratory data. Moreover, our experimental finding is consistent with recent experimental work on reciprocity. Gächter and Thöni (2010) conduct 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 experiments that disadvantageous relative positions negatively affect effort exertion and working performance (see, for example, Clark et al. 2010, Greiner et al. 2011).

## 6. Conclusion

We investigate how bonus payments in a large multinational company, and the degree of bonus transparency,

<sup>33</sup> 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  $> 0.1$ .



affect satisfaction and performance. Utilizing a unique data set that combines survey and compensation data, we find that if a manager's bonus falls behind a reference point, self-reported job satisfaction is significantly reduced; being ahead hardly affects 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 toward the reference point. These effects are mitigated, however, if information about one's relative standing toward the reference point is withheld.

We show that our main observations are predicted by combining straightforward and widely used behavioral models. In particular, we show that uncertainty about one's relative standing toward the reference point reduces the impact of reference-dependent preferences and that adding a framework of subjective performance evaluations to our model explains that bonuses are clustered at the reference point if reference point violations are transparent. In fact, we can rule out several competing explanations for the impact of transparent reference point violations, such as signaling of future career prospects. Moreover, our main observations are supported by a complementary laboratory experiment that controls for potentially confounding influences. Finally, our observations regarding the role of 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 (Mussweiler and Damisch 2008 provide a review of the more recent relevant cognitive psychology research).<sup>34</sup>

One reason why we find quite strong and robust effects of transparent reference point violations may be that the 100% bonus percentage is both a reasonable reference point for social comparisons and a reasonable aspiration for "fully meets expectations" managers. Although the two forces might well reinforce each other, we note that we could not find clear evidence for an aspiration effect, yet we have provided laboratory and field evidence suggesting that social comparison is at least part of the explanation. However, it is an interesting task for future research to further disentangle the contribution of those two potential reference standards.

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 near future" (Towers Perrin 2007, p. 12).<sup>35</sup> However, the effectiveness of many commonly used incentive instruments, and their interaction, is not yet well understood.

From this perspective, the contribution of our study is to show how the choice of transparency affects satisfaction, performance, and evaluations. Yet although our study indicates that transparency comes at a cost,<sup>36</sup> 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 Schaubroeck et al. 1994 and Werner and Ones 2000 for examples in psychology and Bolton et al. 2005 and Frey and Stutzer 2005 for examples in economics). One reason is that a lack of transparency bears the risk of inequitable pay rates or inequitable procedures to allocate bonuses that would be objected to if transparent. Our study indeed illustrates that a less transparent system may exhibit significantly more reference point violations than a more transparent system.<sup>37</sup> Pay secrecy might reduce trust in the organization and increase the fear of employees of being exploited (Colella et al. 2007).

Overall, in light of the arguments in favor of transparent procedures and outcomes, and 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 changed as a reaction to our study. One important change was to fix

<sup>35</sup> This report includes 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.

<sup>36</sup> A related observation has been made by Frank and 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."

<sup>37</sup> Transparency of pay rates and bonuses is also on the political agenda in the United States. For instance, in August 2009, Colorado joined California, Michigan, and Illinois in protecting the right of employees to discuss their wages. The so-called Wage Transparency Act prohibits employers from taking adverse actions against employees who discuss their wages with others.

<sup>34</sup> See also Benjamin et al. (2010) and Chen et al. (2010) for an application in the economics literature.

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.

### Acknowledgments

The authors thank the associate editor, three anonymous referees, and many colleagues who commented on earlier versions of this paper, as well as seminar audiences at Aarhus, Bonn, Cologne, Dortmund, Edinburgh, Granada, Heidelberg, Karlsruhe, London, Magdeburg, Mannheim, Munich, and Toulouse for very helpful feedback that greatly improved our paper. The authors also thank the board and many managers of the studied company for sharing their data and views with us. Finally, the authors thank the German Research Foundation for financial support through the Gottfried Wilhelm Leibniz Program and through the research unit “Design & Behavior” [FOR 1371].

### References

- Akerlof G, Yellen J (1990) The fair wage-effort hypothesis and unemployment. *Quart. J. Econom.* 105(2):255–283.
- Ariely D, Ockenfels A, Roth AE (2005) An experimental analysis of ending rules in internet auctions. *RAND J. Econom.* 36(4):890–907.
- Barankay I (2012) Rank incentives: Evidence from a randomized workplace experiment. Working paper, University of Pennsylvania, Philadelphia.
- Bell DE (1985) Disappointment in decision making under uncertainty. *Oper. Res.* 33(1):1–27.
- Benjamin DJ, James JC, Strickland AJ (2010) Social identity and preferences. *Amer. Econom. Rev.* 100(4):1913–1928.
- Bewley TF (1999) *Why Wages Don't Fall During a Recession* (Harvard University Press, Cambridge, MA).
- Bolton GE (1991) A comparative model of bargaining: Theory and evidence. *Amer. Econom. Rev.* 81(5):1096–1136.
- Bolton GE, Ockenfels A (2000) ERC: A theory of equity, reciprocity and competition. *Amer. Econom. Rev.* 90(1):166–193.
- Bolton GE, Ockenfels A (2010) Betrayal aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States: Comment. *Amer. Econom. Rev.* 100(1):628–633.
- Bolton GE, Brandts J, Ockenfels A (2005) Fair procedures: Evidence from games involving lotteries. *Econom. J.* 115(506):1054–1076.
- Brandts J, Saijo T, Schram A (2004) How universal is behavior? A four country comparison of spite, cooperation and errors in voluntary contribution mechanisms. *Public Choice* 119(3–4):381–424.
- Card D, Mas A, Moretti E, Saez E (2012) Inequality at work: The effect of peer salaries on job satisfaction. *Amer. Econom. Rev.* 102(6):2981–3003.
- Chen YF, Harper M, Konstan J, Li SX (2010) Social comparisons and contributions to online communities: A field experiment on MovieLens. *Amer. Econom. Rev.* 100(4):1358–1398.
- Clark AE (1999) Are wages habit-forming? Evidence from micro data. *J. Econom. Behav. Organ.* 39(2):179–200.
- Clark AE, Oswald AJ (1996) Satisfaction and comparison income. *J. Public Econom.* 61(3):359–381.
- Clark AE, Frijters P, Shields MA (2008) Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles. *J. Econom. Literature* 46(1):95–144.
- Clark AE, Masclet D, Villeval MC (2010) Effort and comparison income: Experimental and survey evidence. *Indust. Labor Relations Rev.* 63(3):407–426.
- Cohn A, Fehr E, Herrmann B, Schneider F (2014) Social comparison and effort provision. *J. Eur. Econom. Assoc.* 12(4):877–898.
- Colella A, Paetzold R, Zardkoohi A, Wesson M (2007) Exposing pay secrecy. *Acad. Management Rev.* 32(1):55–71.
- Cooper DJ, Kagel JH (2014) Other regarding preferences: A selective survey of experimental results. Kagel JH, Roth AE, eds. *The Handbook of Experimental Economics*, Vol. 2 (Princeton University Press, Princeton, NJ). Forthcoming.
- Demougin D, Fluet C (2003) Inequity aversion in tournaments. CIRPEE Working Paper 03-22, Centre interuniversitaire sur le risque, les politiques économiques et l'emploi, Montreal.
- Dufwenberg M, Kirchsteiger G (2004) A theory of sequential reciprocity. *Games Econom. Behav.* 47(2):268–298.
- Falk A, Fischbacher U (2006) A theory of reciprocity. *Games Econom. Behav.* 54(2):293–315.
- Fehr E, Schmidt K (1999) A theory of fairness, competition, and cooperation. *Quart. J. Econom.* 114(3):817–868.
- Fehr E, Götte L, Zehnder C (2009) A behavioral account of the labor market: The role of fairness concerns. *Annual Rev. Econom.* 1:355–384.
- Fehr E, Kirchsteiger G, Riedl A (1993) Does fairness prevent market clearing? An experimental investigation. *Quart. J. Econom.* 108(2):437–460.
- Ferrer-i-Carbonell I (2005) Income and well-being: An empirical analysis of the comparison income effect. *J. Public Econom.* 89(5–6):997–1019.
- Fischbacher U (2007) z-Tree: Zurich toolbox for ready-made economic experiments. *Experiment. Econom.* 10(2):171–178.
- Frank RH (1984) Are workers paid their marginal products? *Amer. Econom. Rev.* 74(4):549–571.
- Frank RH (1985) The demand for unobservable and other nonpositional goods. *Amer. Econom. Rev.* 75(1):101–116.
- Frank RH, Sunstein CR (2001) Cost-benefit analysis and relative position. *Univ. Chicago Law Rev.* 68(2):323–374.
- Frey BS, Stutzer A (2002) What can economists learn from happiness research? *J. Econom. Literature* 40(2):402–435.
- Frey BS, Stutzer A (2005) Beyond outcomes: Measuring procedural utility. *Oxford Econom. Papers* 57(1):90–111.
- Freeman RB (1978) Job satisfaction as an economic variable. *Amer. Econom. Rev.* 68(2):135–141.
- Gächter S, Fehr E (2002) Fairness in the labour market: A survey of experimental results. Bolle F, Lehmann-Waffenschmidt M, eds. *Surveys in Experimental Economics: Bargaining, Cooperation and Election Stock Markets* (Physica, Heidelberg, Germany), 95–132.
- Gächter S, Thöni C (2010) Social comparison and performance: Experimental evidence on the fair wage-effort hypothesis. *J. Econom. Behav. Organ.* 76(3):531–543.
- Gartenberg C, Wulf J (2013) Pay harmony: Peer comparison and executive compensation. Working paper, New York University, New York.
- Greiner B (2004) An online recruitment system for economic experiments. Kremer K, Macho V, eds. *Forschung und wissenschaftliches Rechnen, GWDG Bericht 63* (Gesellschaft für wissenschaftliche Datenverarbeitung, Göttingen, Germany), 79–93.
- Greiner B, Ockenfels A, Werner P (2011) Wage transparency and performance: A real-effort experiment. *Econom. Lett.* 111(3):236–238.
- Grund C, Sliwka D (2005) Envy and compassion in tournaments. *J. Econom. Management Strategy* 14(1):187–207.

- Hamermesh D (1977) Economic aspects of job satisfaction. Ashenfelter OC, Wallace EO, eds. *Essays in Labor Market and Population Analysis* (John Wiley & Sons, New York), 53–72.
- Hamermesh D (2001) The changing distribution of job satisfaction. *J. Human Resources* 36(1):1–30.
- Harrison GW, List JA (2004) Field experiments. *J. Econom. Literature* 42(4):1009–1055.
- Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–291.
- Kagel JH, Roth AE (2000) The dynamics of reorganization in matching markets: A laboratory experiment motivated by a natural experiment. *Quart. J. Econom.* 105(1):201–235.
- Kőszegi B, Rabin M (2006) A model of reference-dependent preferences. *Quart. J. Econom.* 121(4):1133–1165.
- Kwon I, Meyersson Milgrom E (2009) Status in the workplace: Evidence from M&A. Working paper, Seoul National University, Seoul, Korea.
- Lazear EP, Rosen S (1981) Rank-order tournaments as optimum labor contracts. *J. Political Econom.* 89(5):841–864.
- Levine DI (1993) What do wages buy? *Admin. Sci. Quart.* 38(3): 462–483.
- Liu C, Borg I, Spector PE (2004) Measurement equivalence of the German job satisfaction survey used in a multinational organization: Implications of Schwartz’s culture model. *J. Appl. Psych.* 89(6):1070–1082.
- Loomes G, Sugden R (1986) Disappointment and dynamic consistency in choice under uncertainty. *Rev. Econom. Stud.* 53(2): 271–282.
- Luttmer E (2005) Neighbors as negatives: Relative earnings and well-being. *Quart. J. Econom.* 120(3):963–1002.
- Mas A (2006) Pay, reference points, and police performance. *Quart. J. Econom.* 121(3):783–821.
- Mussweiler T, Damisch L (2008) Going back to Donald—How comparisons shape judgmental priming effects. *J. Personality Soc. Psych.* 95(6):1295–1315.
- Nagin DS, Rebitzer JB, Sanders S, Taylor LJ (2002) Monitoring, motivation, and management: The determinants of opportunistic behavior in a field experiment. *Amer. Econom. Rev.* 92(4):850–873.
- Neckermann S, Cueni R, Frey BS (2014) Awards at work. *Labor Econom.* 31:205–217.
- Oosterbeek H, Sloof R, Van De Kuilen G (2004) Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Experiment. Econom.* 7(2):171–188.
- Pfeffer J, Langton N (1993) The effect of wage dispersion on satisfaction, productivity, and working collaboratively: Evidence from college and university faculty. *Admin. Sci. Quart.* 38(3):382–407.
- Prendergast C (2002) Uncertainty and incentives. *J. Labor Econom.* 20(2):115–137.
- Prendergast C, Topel RH (1996) Favoritism in organizations. *J. Political Econom.* 104(5):958–978.
- Rabin M (1993) Incorporating fairness into game theory and economics. *Amer. Econom. Rev.* 83(5):1281–1302.
- Rebitzer JB, Taylor LJ (2011) Extrinsic rewards and intrinsic motives: Standard and behavioral approaches to agency and labor markets. Ashenfelter O, Card D, eds. *Handbook of Labor Economics*, Volume 4A (Elsevier, North-Holland, Amsterdam), 701–772.
- Schaubroeck J, May DR, Brown FW (1994) Procedural justice explanations and employee reactions to economic hardship: A field experiment. *J. Appl. Psych.* 79(3):455–460.
- Towers Perrin (2007) Reward management: Closing a growing say/do gap. Reward challenges and changes survey report. Towers Perrin, Stamford, CT.
- Werner S, Ones DS (2000) Determinants of perceived pay inequities: The effects of comparison other characteristics and pay-system communication. *J. Appl. Soc. Psych.* 30(6):1281–1309.