Go beyond (your) average: A field experiment on real-time performance feedback and sales productivity*

Angela Steffen† Frauke von Bieberstein‡

Working Paper, July 17, 2018

Abstract

Real-time performance feedback is one of the major trends in human resource management. However, insights about the benefits of real-time feedback, i.e. the frequent and immediate disclosure of performance-related information, are still scarce. We present results of a RCT run with a large Railway catering company in Switzerland. In the presence of a relative incentive scheme, we find that real-time information about average performance levels can significantly increase sales productivity. Using a mixed effects regression, we observe a revenue growth of 3-4%, depending on the specification, which corresponds to over 0.4 Million Swiss Francs per year. The effect is mainly driven by employees performing just below the average productivity level, whereas the top- and poorest-performing workers do not show significant reactions.

Keywords: relative performance feedback, field experiment, employee motivation, real-time feedback

JEL Codes: D02, M59, C93

*We thank the management of the partner company who made this project possible. We are grateful to Christian Zehnder, Holger Herz and the participants of the workshop on feedback and recognition at the Erasmus University Rotterdam, the CUSO workshop and the internal research seminar at the University of Bern. All errors are ours.

†angela.steffen@iop.unibe.ch, Institute for Organization and Human Resource Management, University of Bern

‡frauke.vonbieberstein@iop.unibe.ch, Institute for Organization and Human Resource Management, University of Bern

1 Introduction

Organizations are radically changing the way they measure, evaluate, and recognize employee performance. PricewaterhouseCoopers reports for example that two-thirds of large companies in the UK are in the process of adapting their performance management system (Deloitte University Press, 2015). According to Deloitte (2015), 82% of the surveyed U.S. companies perceive traditional performance evaluations as not being worth the time. Currently, the amount of data collected and the speed with which performance-related information can be made available to employees and managers is rapidly expanding. A related trend in the present "performance management revolution" is the shift from year-end appraisals towards a continuous feedback culture with real-time reviews of employee's performance (Cappelli & Tavis, 2016; Deloitte University Press, 2015; The Economist, 2016). Goldman Sachs and JP Morgan are just two recent examples (Son, 2017-03-09; Surane, 2017-04-21).

Yet, in theory and managerial practice, there is still much uncertainty about what constitutes effective performance feedback. Existing evidence supports the idea that immediate performance information leads to better performance than delayed feedback (e.g. Fajfar, Campitelli, & Labollita, 2012; Kettle & Häubl, 2010) and that more specific feedback should be more beneficial (e.g. Casas-Arce, Lourenço, & Martínez-Jerez, 2017; Earley, Northcraft, Lee, & Lituchy, 1990). Furthermore, literature on feedback frequency generally suggests a positive effect of providing more regular performance information (Kang, Oah, & Dickinson, 2005; Kuhnen & Tymula, 2012; Northcraft, Schmidt, & Ashford, 2011). Other studies, however, find opposite results (e.g. Casas-Arce et al., 2017; Lurie & Swaminathan, 2009). The effect of real-time feedback, characterized by the automated provision of frequent and immediate performance information, and its impact over time remain mostly unexplored.

In this study, we investigate the effect of real-time feedback in a real work setting. In particular, we compare different types of real-time feedback to a coarse, monthly performance

signal, which is the basis for relative monetary rewards at the end of the month. Insights about effective feedback provision in the presence of relative rewards are of particular interest, as relative incentive schemes are highly pervasive in practice and employees inevitably receive information on their relative standing in these settings (at the latest when bonuses are distributed at the end of the evaluation period). We conduct a field experiment in a large Swiss catering enterprise with around 200 sales employees offering drinks and snacks on domestic trains. In three experimental treatments, employees frequently receive personal and/or co-worker-related performance information directly at work. This information is given in addition to a coarse performance summary which is only provided ex-post to the monthly bonus payment.

The performance messages in our seeting either contain the employee's personal average sales revenue over the recent past (personal performance information), the recent average sales revenue of all employees (social performance information) or both. Prior economic papers have repeatedly demonstrated that the provision of relative rank information, where individuals learn their relative standing compared to their peers, can increase performance, even when outcomes are not tied to pecuniary rewards (Azmat & Iriberri, 2016; Blanes i Vidal & Nossol, 2011; Delfgaauw, Dur, Sol, & Verbeke, 2013; Kuhnen & Tymula, 2012; Tran & Zeckhauser, 2012). Other studies, however, report negative results of relative performance feedback (e.g. Bandiera, Barankay, & Rasul, 2013; Barankay, 2011; Eriksson, Poulsen, & Villeval, 2009). Social psychology research has further shown that mere information about personal past performance levels or social performance standards can effectively increase effort (Schultz, 1999; Schultz, Nolan, Cialdini, Goldstein, & Griskevicius, 2007; Szymanski & Harkins, 1987; White, Kjelgaard, & Harkins, 1995).

We find that frequent real-time feedback about co-workers' average performance levels leads to a strong productivity increase. Furthermore, this motivational impact seems to be stable over time, offering substantial economic benefits. A further investigation of employees

at different ability levels reveals that the effect of real-time feedback is not uniform: In line with existing evidence on relative performance feedback (Bandiera et al., 2013; Casas-Arce & Martínez-Jerez, 2009; Delfgaauw, Dur, Non, & Verbeke, 2014; Hannan, Krishnan, & Newman, 2008), the positive results in our study are driven by employees at intermediate performance levels, particularly by those who usually perform just below the average. Top-performers, on the other hand, are not affected, which we explain by a ceiling effect. Workers at the bottom of the productivity distribution even show slightly negative reactions. While we observe a similar behavioral pattern for personal real-time feedback, the mere provision of personal performance averages has no significant effect on revenues.

In addition to the comparison of different feedback contents within a relative incentive scheme, the main novelty of our design is that performance information is provided on an ongoing basis, directly at work. The four month duration of our study also allows to investigate the impact of real-time feeback over time. Furthermore, we here have the opportunity to address a rather low-educated workforce, which has been a less investigated target group in the existing feedback literature.

The remainder of this paper is set out as follows: Section 2 provides an overview over related literature and research streams. The field setting and experimental design of our study are described in Section 3. Section 4 shows the derived hypotheses. The data structure, regression models and findings are presented in Section 5. Section 6 eventually provides a discussion of the results and approaches for future research.

2 Related literature

2.1 Comparative feedback effects

Performance feedback has been studied extensively in the field of organizational behavior management and used successfully to increase performance in a variety of organizational settings (see Nolan, Jarema, & Austin, 1999). One main argument for the explanation of feedback effects is that behavior is regulated by comparisons of performance outcomes to goals or standards [1] (Kluger & DeNisi, 1996, p. 259). Existing literature has therefore repeatedly emphasized the importance of self-evaluation for lasting motivational impacts (Bandura & Cervone, 1983; Ilgen, Fisher, & Taylor, 1979; Locke et al., 1981; Strang, Lawrence, & Fowler, 1978). Alvero, Bucklin, and Austin (2001, p. 19) accordingly identified two types of feedback inventions as being equally popular in the feedback literature: The comparison of an individual's performance to his or her past performance (temporal comparative information), and the comparison of individual performance with a standard or mean of performance (social comparative information).

The fact that people are influenced by temporal or social comparative information has been repeatedly documented in social psychology and economics. In the context of socially desirable behavior, social standards and the related opportunity for self-evaluation were proven to be an effective mean to overcome social loafing and improve individual performance in different settings (see Buunk & Gibbons, 2007). By providing information about the average performance of their peers, individuals for example increased performance in a brainstorming task (Szymanski & Harkins, 1987; White et al., 1995), enhanced curbside recycling (Schultz, 1999) or reduced household energy consumption(Schultz et al., 2007). In similar vein, self-based standards from past performance were shown to have a significant

[1] Locke, Shaw, Saari, and Latham (1981, p. 126) and Latham and Locke (1991, p. 231) describe goals as the "aim or end of an action" which also constitutes a standard towards which people evaluate their performance. Standards are defined as "a rule to measure or evaluate things."

influence on behavior, when they are activated through feedback. In a field experiments on recycling behaviour, Schultz (1999) found that individual feedback about the current and previous amount of material a household has collected, significantly improved curbside recycling. Other studies have shown that even the communication of simple attainment levels (also defined as "knowledge of results") can improve performance in real work settings. In his study with low-paid workers executing a repetitive, industrial task, Hundal (1969) found a significant performance improvement as the extent and accuracy of information about personal performance levels increased. These findings are consistent with various other field studies (Braunstein, Klein, & Pachla, 1973; Crowell, Anderson, Abel, & Sergio, 1988; Kim & Hamner, 1976; Nordstrom, Hall, Lorenzi, & Delquadri, 1988; Sharma, Chevidikunnan, Khan, & Gaowgzeh, 2016).

Recent economic literature has particularly focused on feedback containing relative performance information such as ranks or other indicators of an individual's relative position within his or her peer group. Many studies conclude that potential disadvantages of revealing relative performance information to employees are outweighed by its benefits. In a field-experiment with German warehouse employees Blanes i Vidal and Nossol (2011) show that relative performance feedback leads to a large and lasting productivity increase. The workers in their study are paid piece-rates and receive information about their relative position in the payment and productivity distribution. In a field experiment with a Dutch retail chain Delfgaauw et al. (2013, pp. 315-316) also find that tournaments without monetary rewards, but relative performance feedback solely, have a positive effect on sales growth. The feedback effect in their study is of similar magnitude as the effect of tournaments with monetary rewards, suggesting a high symbolic value of winning a tournament. Kuhnen and Tymula (2012) show in a laboratory experiment that relative rank feedback creates a ratcheting effect in productivity, also in the absence of an explicit tournament and without any kind of performance-related pay. The output increase in their setting can be mainly traced back to the fight for dominance among top performers. In the school context, Tran and

Zeckhauser (2012) and Azmat and Iriberri (2010) find strong positive effects on test grades, if students learn their relative standing in previously taken exams.

On the other hand, some studies have reported negative effects of relative performance feedback. In a field experiment with fruit pickers, Bandiera et al. (2013) find that daily histograms on team's productivity reduce overall performance, if relative productivity is not tied to monetary rewards and therefore leads to excessive free-riding within teams. On an individual level, Barankay (2011) demonstrates in another field study with piece-rate pay that relative rank information reduces the productivity of workers as well as the probability that they return to work. Eriksson et al. (2009) show in a lab experiment that relative performance feedback does not improve performance, regardless of a piece-rate or tournament payment scheme.[2] At the same time, feedback information leads to a negative "quality peer effect" because the work-quality of poorly performing employees is reduced. Akın and Karagözoğlu (2017) also demonstrate a negative impact of relative performance feedback under piece-rate incentives, however, they explain this result by the combination of an intense task and frequent feedback messages, causing distraction in their setting.

By further investigating the contradictory evidence, several studies have underlined that the effect of relative performance feedback depends on the incentive scheme. In a lab experiment, Hannan et al. (2008) confirm the positive effect of relative performance feedback under piece-rates, but find a detrimental impact under tournament-based incentives, if the feedback is sufficiently precise. Azmat and Iriberri (2016) find a strong performance increase in a real-effort task when participants learn the average task performance of all subjects in addition to their absolute output. However, this kind of feedback increases performance only under piece-rates and has no effects under a flat-rate compensation scheme. Taken together, existing findings indicate that relative performance information is probably most effective when performance is related to pay and when these rewards are rather tied to absolute,

[2]According to the authors, this result might be attributed to subjects exerting maximum effort given their ability.

instead of relative performance.

2.2 Feedback and ability

Another explanation for the varying outcomes of relative performance feedback can be found in its presumed non-linear impact on workers of different ability levels. Existing studies have shown formally and empirically that relative incentive schemes (and related performance feedback about the participant's relative standing during the competition) have a hump-shaped effect on performance: In a contest, participants who lag far behind as well as participants who are far ahead slack off, whereas incentive salience and responsiveness is high for participants at intermediate performance levels. This pattern is also called "dynamic incentive effect" (Bartel, 2004; Casas-Arce & Martínez-Jerez, 2009; Delfgaauw et al., 2014). Under tournament-based incentives, interim, relative performance feedback can therefore deteriorate performance as low-performing participants, who lag far behind, as well as leading participants decrease their effort (see Casas-Arce & Martínez-Jerez, 2009; Hannan et al., 2008). From a social psychology perspective, this finding is in line with the broadly supported proposition that the standards towards which performance is evaluated should be perceived as both, difficult and rewarding (see Campion & Lord, 1982; Tosi, Locke, & Latham, 1991).

Heterogeneous performance effects were also found outside of relative monetary rewards, where, from a pure rational perspective, the feedback sign should not affect performance. Kuhnen and Tymula (2012) for example show that people's beliefs about their relative ability in the reference group have implications for their effort choices – an effect which they relate to self-confidence and self-esteem. In face of relative performance feedback without monetary consequences, agents basically split into two segments: Top performers, who keep fighting for high ranks and bottom performers, who compete much less. In similar vein, social cognitive theory has shown that motivational responses to performance feedback are regulated by perceived self-efficacy, i.e. individuals' beliefs about their capabilities to perform actions at

a certain level of performance (Bandura & Cervone, 1983, 1986; Bandura & Jourden, 1991; Schunk & Swartz, 1993). Bartol, Durham, and Poon (2001) for example demonstrates that the effects of performance ratings on performance improvement are partially mediated by perceived self-efficacy and personal goals. Comparative information like personal progress or relative standing affect motivation by influencing individuals' perceived capabilities to attain certain standards. This leads to a curvilinear relationship between performance-standard discrepancies and subsequent effort (also see expectancy-valence theory, e.g. Feather, 1982; Heckhausen, 1977).

Further empirical studies support this idea. Using quantile regression, Bandiera et al. (2013) show that, given piece-rate-incentives, the introduction of rank feedback reduces the productivity of teams at the bottom of the conditional productivity distribution. On the other hand, it has no effect on teams above the 40th percentile. In a laboratory experiment, Eriksson et al. (2009) confirms that the average score of the worst performers is reduced, when they are continuously updated about their opponent's performance. This effect appears independently of the payment condition (piece-rate or tournament) and can be attributed to an increase in mistakes by low-performing subjects. The average score of the best performers, on the other hand, is not significantly affected. For above average performers, negative feedback effects are particularly reported from norm theory. Schultz et al. (2007) show for example an undesired boomerang effect of peer-related, normative messages for households who were already consuming at a low rate. This finding is consistent with other studies on social performance information (see Fischer, 2008). Azmat and Iriberri (2016) do not find heterogeneous feedback effects on performance, but they show that receiving positive (negative) feedback has a strong positive (negative) impact on participant's emotional states, i.e. their happiness and dominance levels.[3] This result is in line with prior evidence suggesting that people have powerful affective responses to being worse than others (Klein, 1997; Moore

[3]The effect occurs under piece-rate incentives, when relative performance information has consequences in terms of relative income. Azmat and Iriberri (2016) find no impact on satisfaction in a flat-rate pay condition.

& Klein, 2008).

2.3 Feedback frequency and immediacy

In contrast to the introduction of new feedback policies, the specific impact feedback immediacy and feedback frequency has received less attention. While it is generally argued that feedback is most effective when provided on a regular basis, immediately after behavior, some studies relate more frequent feedback to negatives results. In a field experiment with an insurance repair company, Casas-Arce et al. (2017) find that professionals achieve higher customer satisfaction scores when they receive detailed, but infrequent (monthly) feedback. This effect arises because in the frequent feedback condition, previous information is disregarded in the face of new information. Likewise, Lurie and Swaminathan (2009) find in a lab experiment that more regular feedback on previous decisions leads to declines in performance, as decision makers overweight the most recent data received. Chhokar and Wallin (1984) find no effect of more frequent feedback on performance effectiveness in an industrial safety setting.

On the other hand, numerous studies reveal positive outcomes of increasing the frequency or immediacy of feedback – in particular when both features are present. The findings of Kang et al. (2005) for example indicate that more frequent feedback produces a higher level of performance than less frequent feedback when individuals receive incentive payments. In similar vein, the results of So, Lee, and Oah (2013) suggest that more frequent feedback is more effective for improving customer service behaviour. In their within-subject design, employees of a gas station showed small but consistent improvements in service performance when they received daily, as compared to weekly feedback. Northcraft et al. (2011) report a positive impact of more frequent and more specific feedback (a feature which often comes along with more direct or immediate feedback) on performance. The positive effects of timely and specific feedback in their study were accentuated when both characteristics were

combined. Alavosius and Sulzer-Azaroff (1986) and Mason and Redmon (2008) further support a positive impact of feedback immediacy on the acquisition of desired behavior (Alavosius & Sulzer-Azaroff, 1986) and work performance (Mason & Redmon, 2008). Closely related to our study, Goomas, Smith, and Ludwig (2011) investigate the impact of frequent and immediate feedback in a field experiment with order selectors in a warehouse distribution center. Their results indicate that immediate, real-time information about workload and team performance, i.e. comparisons to engineerd labor standards, has a substantial impact on worker's productivity. In other studies, Goomas found similar positive effects of real-time performance goals and feedback in the context of distribution centres in additional studies (Goomas, 2012; Goomas & Ludwig, 2007; **?**).

3 Methodology

3.1 Company setting

Our project partner is a big railway catering enterprise in Switzerland. The largest company unit includes the service of food, snacks and drinks on Swiss trains by so-called 'stewards'. The company employs around 200 'minibar stewards' who sell drinks and snacks from a mobile vending cart, and about 330 'restaurant stewards' who serve customers in the train restaurants. Target group of our experiment are the minibar stewards. In their role as sales-people, rather than service personnel, they have a strong and direct influence on sales performance, for example through their walking speed, friendliness, verbal promotion or cross-selling efforts.[4] Besides the generation of revenue, the steward's motivation and effort also plays a crucial role with regard to customer satisfaction and the company's reputation.

[4]This assessment did not only emerge from interviews with the partner company but is also reflected in the data: The variance partition coefficient which compares the between-employee performance variation to the overall performance variation in the data is 21% for the minibar stewards and 11% for the restaurant stewards.

At the same time, employee motivation is one of the organization's major challenges: The job of the minibar stewards is not highly regarded, rather isolated and repetitive. Furthermore, the work is physically demanding, which also explains to some extent why 98% of minibar stewards are male. Another management challenge is the existing lack of control mechanisms. As minibar stewards usually start, execute and terminate their services alone, there hardly exists any interaction with superiors or co-workers and corresponding controls.

To foster motivation, the company currently uses an incentive scheme, depending on a steward's relative sales performance. This system offers the prospect of significant bonus payments, which - according to the company - account for up to 20% of a worker's monthly income. As revenues highly depend on train routes and service times, the incentive scheme compares sales performance within the same work shifts. A shift starts and ends at a certain time at a certain destination (usually the steward's official work place) and covers a specific train route. There exist between 100 and 200 different minibar shifts, which are usually operated on a daily basis. The stewards work on various shifts, in accordance to the monthly shift plan. At the end of every month, the personal average revenue of a steward on a certain shift is compared to the total average revenue of all stewards who have worked on the same shift.[5] It is further not assumed that they collude in any way, because relational ties among the stewards are very loose and knowledge about co-workers performance is hardly available.By this approach, the company aims to provide an employee evaluation, which is as fair as possible.[6] The weighted mean of the shift-comparisons defines a steward's total performance in a certain month (overall deviation to average shift revenues in %). Currently, the performance differences thus calculated are quite large: the variation between minibar stewards reaches from -22 up to more than 23% with a standard deviation of approximately 11.7%-points (see Appendix A). While there also exists performance variation

[5]Stewards cannot directly learn "their competitors" on a certain shift from the work plan.

[6]To reduce the impact of extraordinary events or happenings, this performance is only calculated for stewards who have worked on at least 11 services in the particular month. Extraordinary events such as train failures are also considered on an individual basis. However, the performance evaluation does not account for other circumstances affecting performance like weekdays or number of passengers on the train.

within employees over time (the average standard deviation of a steward's performance over the months lies at 7.9%-points), between-worker differences are stronger.

Based on the monthly performance measure, the bonus pot is distributed as illustrated in Figure 1. For above-average performance, stewards receive a proportional bonus payment, while stewards below the total average receive no reward. This approach is similar to the proportional prize contest, introduced by Cason, Masters, and Sheremeta (2010), in which a prize is divided among participants in proportion to their achievement. At the end of the month, stewards are informed about their overall performance and the corresponding reward on their salary statement. Apart from this generic overview, stewards hitherto received no comparative performance information such as performance benchmarks or regular feedback from superiors. Our study was designed to exploit this motivational potential, in accordance with the existing incentive scheme.

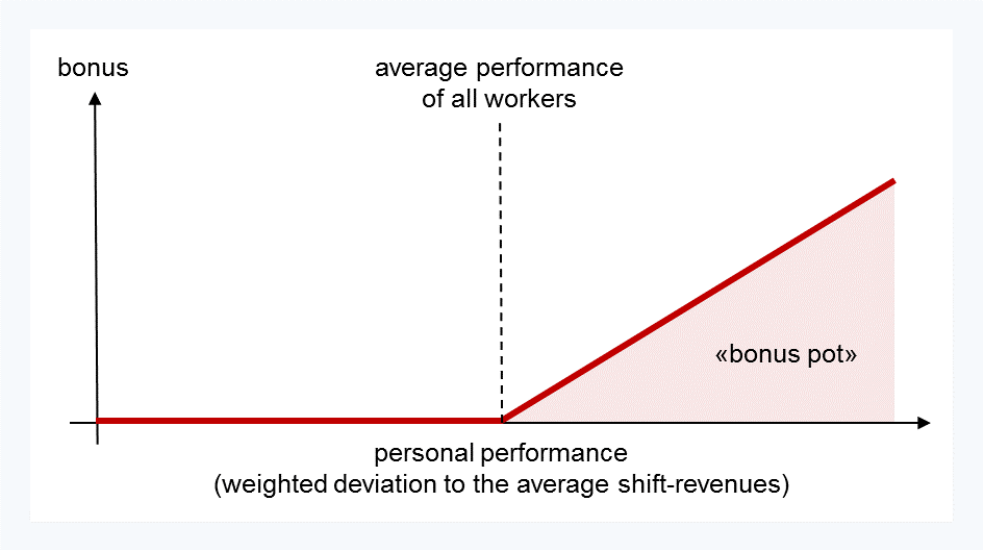

Figure 1: Incentive scheme

3.2 Experimental design

We used a between-subject design consisting of three treatments and one control group. All treatments received ongoing, real-time feedback about recent performance averages of their current work task, which automatically appeared on the electronic cash desk. Group 1 ("personal information") was informed about their own average revenue on the present shift during the last 30-days.[7] Group 2 ("social information") was shown the total average revenue of all stewards working on the same shift during the last 30 days. The message of group 3 (personal & social information) contained both, the shift-specific, average r of all workers as well as the steward's personal average revenue during the past 30 days. This particularly affects our first treatment.Figure 2 provides an example. The control group ("no information") only received a general thank you-message without any performance information.

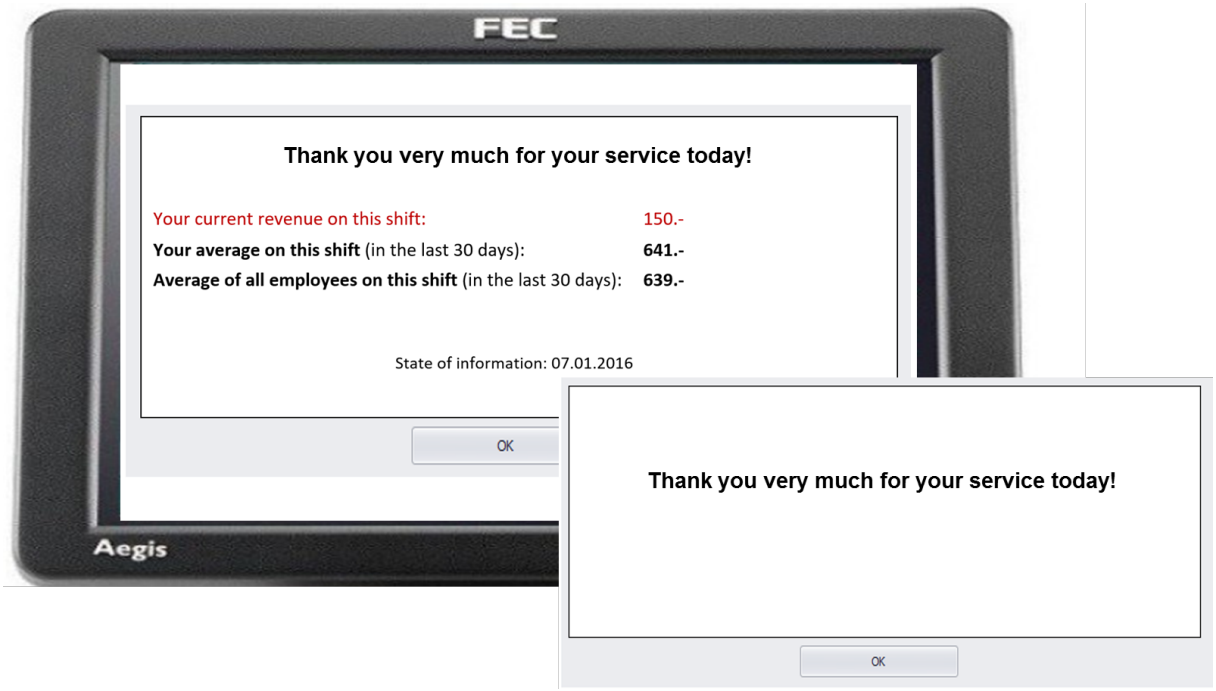

Figure 2: Message example for treatment 3 and control

Recall that the feedback of treatment 3 is a similar, but more frequent and more im-

[7] If a steward did not perform the same shift in the recent past, the message still appeared but with an empty space. These observations were not considered in later analyses.

mediate / task-specific information to what stewards receive in the monthly bonus report. Stewards in neither treatment can clearly infer monetary rewards from the feedback messages. In further contrast to the incentive scheme, we did not provide information about the average revenue per shift in the same month. Reason is that the within-month-comparison would have generated many empty or unreliable messages at the beginning of the month, where the number of services performed on a specific shift is still small. We wanted to keep the informative value of the messages constant over time and therefore chose a dynamic, 30-day timeframe.[8]

The messages of the treatment groups also contained a steward's hitherto generated revenue on his current service. Contrary to the performance averages, this information is accessible on the electronic tills at any time. Stewards especially see the total generated revenue of their service when they do the daily accounting at the end of the day. As in most real work contexts, feedback containing personal performance benchmarks is therefore less novel and can more easily tracked by the employees themselves. This particularly affects our first treatment.

All messages were programmed by an external IT company, which also maintains the electronic till-system of our partner firm. The average individual and total sales revenues were automatically calculated in real-time when the steward logged on at the beginning of every service. The respective performance information appeared on the till screen at three different times per day: at the beginning of the service (login), at the end of the service (logoff) and once at a random time during work (at least 2 hours after start and by latest 1 hour before service end; **see section....**). The general message of the control group, on the other hand, only appeared once at the beginning of the shift. To ensure that stewards read the message, they had to click the "OK"-button before they could proceed with another till-

[8]Another option considered was to offer comparative revenue information relating to the same month in the previous year. As product prices and sales vary on a more frequent basis, we decided to use more timely revenue information, which is also stands in closer connection to bonus calculation. Furthermore, we considered this approach as more consistent with the main idea of real-time feedback.

transaction. Furthermore, language was adapted automatically, depending on the stewards' personnel data (German, French or Italian).

The intervention ran from March 1st to June 30, 2016. The experiment comprised all stewards who were employed at start time and all work shifts, with the exception of extra or charter shifts. Excluded from the study were also foreign train connections, operated by (TGV-)Lyria and SNCF Voyages Italia, as these shifts have different service processes. Stewards were randomly assigned to one of the four experimental groups. By stratification, we ensured a balanced distribution of the different divisions and the stewards' prior sales performance[9] among the groups. Prior to launch, all stewards were informed by a generic message, also transmitted via the electronic cash desk, that the head office would use the tills more frequently as an alternative communication channel. The approximately ten sales managers (direct superiors) received a general information about the provision of additional revenue information, which could vary during an initial test period of the project.

4 Hypotheses

Prior studies have revealed that more frequent and immediate performance feedback usually promotes the effectiveness of feedback interventions. We therefore expect an overall positive impact of real-time feedback on subsequent performance for all types of content. This preposition is also supported by certain particularities of our setting: Feedback is provided via a computer screen, a characteristic which was found to be preferable to having supervisors deliver the feedback personally (Kluger & DeNisi, 1996). Kluger and DeNisi (1996, p. 269) further proposed that the effect of feedback on performance is more positive, the fewer cognitive resources are needed for task performance. This finding is in favour of our design, which involves a repetitive sales task with relatively low cognitive demands.

[9] As no further data was available at that point of time, prior performance was based on the bonus calculations of November and December 2015.

Hypothesis 1 *Real-time feedback about personal and/or social performance increases sales productivity.*

As mentioned earlier, evidence regarding the differential effects of personal versus social-comparative feedback is limited (Moore & Klein, 2008, p. 61). Moore and Klein (2008) suggest that information about one's absolute standing may be more influential than social comparative feedback. Klein (1997), on the other hand, demonstrates in the area of health-care risks that manipulations of social risk information has a greater impact than personal risk information on a variety of outcomes including, emotions, intentions and behaviour across several situations. Closer related to our study, Blader, Gartenberg, and Prat (2015) show in a field experiment with truck drivers that providing rank information with respect to other drivers leads to better performance, than informing the drivers about their individual performance solely. In our study, we equally expect the effect of social-related feedback to be stronger than personal feedback (and we expect an even stronger impact for the social and personal feedback condition). This assumption is also based on the fact that in our setting, social performance information is more directly linked to monetary incentives. Previous work has confirmed that feedback combined with consequences produces more consistent effects than feedback alone (see Alvero et al., 2001). Furthermore, co-worker-related performance information in our design is more novel as compared to personal performance averages, which are theoretically trackable by the stewards themselves.

Hypothesis 2 *The effect on sales productivity is highest for personal and social information and lowest for personal performance information alone.*

Finally, despite a presumed positive impact of real-time feedback, we do not expect the treatment effects in our study to be homogenous. Literature around feedback and ability has demonstrated for a broad range of incentive schemes that relative performance information can have a negative influence on workers at the bottom and at the top of the productivity

distribution. Incentives to increase output were, however, shown to be particularly strong for average performers. By making a person's relative standing more salient, we expect that real-time feedback containing social performance information is more effective for workers at intermediate levels of ability as compared to low-performing and high-performing employees. This holds in particular for the personal and social information treatment, where performance discrepancies between an individual worker and his peer group are most salient. On the other hand, we do not presume such heterogeneous effects if employees receive real-time feedback about their personal average performance only. The performance-standard discrepancies in this treatment and the related psychological and monetary consequence for the bottom and top-performing employees are presumably much lower.

Hypothesis 3 *The effect of social performance information and, even stronger, the effect of personal and social performance information is higher for employees at intermediate levels of performance than for workers who usually perform at the extremes.*

5 Results

5.1 Field data and empirical strategy

Our dataset consists of approximately 33'000 minibar service observations. These are all minibar shifts performed by the minibar stewards between January 1st 2015 and end of June 2016. Table table:sample provides an outline of the number of observations per treatment and the main sample characteristics of the cleaned data set (see Section 5.2).

Table 1: Sample Characteristics

	Personal Info	Social Info	Personal+Social	Control	Total
N (services) pre-study	6'265	7'092	7'086	6'441	26'884
N (services) study	1'242	1'902	1'291	1'745	6'180
Number of ID (stewards)	43	44	41	43	171
Data before study period:					
log revenue (CHF)	5.79	5.76	5.81	5.79	5.79
	(0.49)	(0.49)	(0.5)	(0.46)	(0.49)
log revenue per hour (CHF)	3.97	3.93	3.97	3.96	3.96
	(0.41)	(0.41)	(0.41)	(0.39)	(0.4)
log articles per hour	2.46	2.42	2.47	2.45	2.45
	(0.42)	(0.43)	(0.43)	(0.41)	(0.42)
log customers per hour	1.98	1.95	1.98	1.98	1.97
	(0.48)	(0.5)	(0.5)	(0.5)	(0.5)
log articles sold per customer	0.49	0.48	0.49	0.47	0.48
	(0.27)	(0.3)	(0.29)	(0.31)	(0.3)
tenure (years)	7.25	5.17	6.52	7.55	6.58
	(5.94)	(4.62)	(4.76)	(7.33)	(5.79)
workload (ø services per month)	16.71	16.55	16.29	17.06	16.64
	(2.42)	(2.28)	(2.91)	(2.89)	(2.65)
worktime (hours)	6.47	6.58	6.53	6.49	6.52
	(1.92)	(2.14)	(1.96)	(1.77)	(1.96)
break (hours)	1.90	1.67	1.84	1.75	1.79
	(1.3)	(1.17)	(1.24)	(1.21)	(1.23)
train occupancy (%)	37.44	38.33	37.78	37.34	37.74
	(10.8)	(11.45)	(10.84)	(10.44)	(10.91)
share 1. class passengers	18.60	18.51	18.60	18.45	18.54
	(5.59)	(5.77)	(5.66)	(5.47)	(5.63)

Notes: Table shows means of cleaned data (see below) before study period; standard deviations in parentheses

Our main outcome variable (sales performance) is defined as the revenue per hour (log) on each service.[10] A service is defined as a shift, performed by a certain steward on a certain date. Our data therefore includes steward-related (level 2) as well as shift-, date- or service-related characteristics (level 1). We use a logarithmic transformation of the outcome variable

[10]We did not use an aggregated performance measure at the steward-level as dependent variable (e.g. the bonus calculation, see Section 3.1) for two reasons: First, there are major concerns with focusing on only one level in hierarchical data structures. The loss of variance information at any level can lead to a severely incomplete or even misleading knowledge (see Bullen, Jones, & Duncan, 1997; Subramanian, Jones, Kaddour, & Krieger, 2009). This peril is particularly high in our case, where we observe high variation on the lower (service) data level. Secondly, analyzing means on the steward-level makes inference highly volatile. The results of such an analysis strongly depend on the exact specification of the calculated performance measure.

to meet the assumption of normal distribution for parametric statistical tests. To investigate the effect of the experimental treatments on sales performance, we use the following two-level random intercept model, including random effects at the steward level (also see Cameron & Trivedi, 2010, pp. 235ff.). In our regression model, revenue per hour for steward i, shift j and date t is defined as:

$$Ln(revhour)_{ijt} = \beta_0 + \beta_1 Group_i + \beta \boldsymbol{Stew'}_i + \beta \boldsymbol{Shift'}_j + \beta \boldsymbol{Date'}_t + \beta \boldsymbol{Service'}_{ijt} + \upsilon_{0i} + \epsilon_{0ijt} \quad (1)$$

Group is a dummy variable defining our experimental treatment. Stew' is a vector containing steward-specific control variables. These are tenure, employment status (temporary or permanent) and workload (average number or services per month). In order to control for performance ability, we also integrate an indicator of the stewards' average sales performance before our intervention. This measure is computed in the same manner as the monthly bonus calculation of our partner company (i.e. the mean deviation between personal and total shift-averages, see Section 3.1). To calculate the overall pre-study period performance, we took the weighted average of the monthly performance evaluations.

Shift' is a vector containing shift-related variables like type of shift, city of shift start and shift duration (work time). We also created a variable indicating to what extent the shift covers common eating times, i.e. breakfast, lunch and dinner (in % of total work time). Like the shift-vector, Date' also refers to the "first level" of our data and is a vector with time-varying variables influencing consumption on the trains, such as "month" and a dummy variable for weekends or holidays. Furthermore, we control for service-related characteristics Serv' (i.e. information associated with a specific shift performed by a certain steward at a specific date). These controls include service occupancy - a measure for how busy the

service was. Occupancy shows the percentage share of occupied train seats (mean over all trains which were covered by the respective service) and was computed by using the daily passenger numbers of Swiss Federal Railways.[11] We also consider the share of 1st class passengers, which has a strong effect on consumption. In addition, we take into account how many stewards were working on a particular service (in rare cases, there are more than 1) and if the service was affected by a big event taking place nearby the route.

The reason why we control for several shift, date- and service-related variables in addition to occupancy is that they presumably affect consumption patterns beyond the mere amount of passengers. It is for example likely that passengers consume more during weekends or that passenger types and spending behavior vary with respect to the city of shift start. Note that we randomized our treatments only on the steward-level (level 2), but not on the service-level (level 1). The consideration of level 1-controls is therefore crucial.

5.2 Real-time feedback and sales performance

We estimate model (1) with a common multi-level-model command, including robust standard errors. Observations were excluded if passenger data for the specific service was incomplete, if the steward did not accurately receive the performance information of our treatments (e.g. because he did not work on the same shift during the last 30 days) or if the service was affected by a train failure.[12] For the sake of parsimony, we further excluded variables with no significant effects, which were tenure, employment status and workload. Including these

[11] We did not use absolute passenger numbers but occupancy because we want to model a non-linear relationship between the share of occupied seats and sales revenues. We presume lower sales in very crowded trains. Adding passenger numbers in addition to occupancy has no significant effect. The expected non-linear relationship between the number of passengers and sales performance is also the reason why we did not use revenue per passenger as an outcome measure in our regression model.

[12] Using these specifications, we had to exclude 28.5% of the minibar services during the study period. Incomplete performance information and missing passenger data account for the most part of these cases. Performance information was particularly incomplete in treatments 1 and 3. Reason is that stewards did not necessarily work on the same shift during the last 30 days before message release, leading to a missing personal average. 30%, respectively 34%, of all observations in treatment 1 and 3 had to be excluded. In the second treatment (social info), these were only 2%, which explains the lower standard errors in this group.

controls has a negligible influence on the results. Table 2 shows the estimates of model (1) during our intervention. We also looked at additional outcome measures such as the number of articles sold and the number of different customers per hour.

The results show a significant increase in revenue per hour of 3.3% for social information treatment and 3.8% in the personal and social information treatment, compared to the control group. Sales growth is even more strongly reflected in the number of products sold. A closer look at the coefficients in the last two columns indicates that this effect can be mainly attributed to an increase in the number of customers, rather than to enhanced cross-selling activities (additional products per customers).

The results above are stable when conducting various robustness-checks. Appendix B shows the results of the same model for a linear OLS regression with cluster-robust standard errors, which are used to control for heteroscedasticity and an unequal correlation of errors between stewards over time (see Colin Cameron & Miller, 2015). Following the approaches of Fehr and Goette (2007) and Gino and Staats (2011) in a similar setting, we additionally run several difference-in-difference (DID) analyses (see Appendix B). These estimates demonstrate that: 1) We obtain similar results when comparing the pre- and during study period, with significant treatment-effects for the social as well as the personal and social information group, (there are hardly any differences between the treatments before our intervention). 2) The effects also persist when using fixed, instead of random effects at the steward-level and 3) the results in the DID-models are also stable towards adaptions in the control variables and hold without taking account for any controls.

We further tested whether the productivity increase could be attributed to a short-term enhancement of motivation when feedback was launched at the beginning of the study or if the effects persist over time. We find no significant interaction effects between our

Table 2: Log revenues, sales and customers

	log revenue per hour	log articles per hour	log customers per hour	log articles per customer
personal info	0.0151	0.0179	0.0005	0.0131
	(0.0164)	(0.0168)	(0.0192)	(0.0099)
social info	0.0323***	0.0385***	0.0227	0.0162
	(0.0123)	(0.0134)	(0.0163)	(0.0103)
personal + social	0.0390**	0.0500***	0.0364*	0.0113
	(0.0152)	(0.0164)	(0.0195)	(0.0090)
performance before (in %)	0.0101***	0.0101***	0.0100***	0.0000
	(0.0006)	(0.0006)	(0.0007)	(0.0004)
service duration (in h)	-0.0181***	-0.0151***	-0.0252***	0.0105***
	(0.0041)	(0.0045)	(0.0051)	(0.0021)
breakH	0.0191**	0.0296***	-0.0050	0.0349***
	(0.0082)	(0.0086)	(0.0122)	(0.0044)
eating times (in %)	0.0078***	0.0127***	0.0100***	0.0026***
	(0.0006)	(0.0007)	(0.0007)	(0.0003)
occupancy (in %)	0.0235***	0.0234***	0.0212***	0.0022*
	(0.0027)	(0.0028)	(0.0028)	(0.0012)
occupancy^2	-0.0001***	-0.0001***	-0.0001***	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
1st class pass (in %)	0.0096***	0.0110***	0.0124***	-0.0013**
	(0.0012)	(0.0013)	(0.0013)	(0.0006)
No. stewards working	-0.7375***	-0.7356***	-0.5396***	-0.0384
	(0.1898)	(0.1697)	(0.0915)	(0.0464)
event	0.1467***	0.1374***	0.0949***	0.0408***
	(0.0304)	(0.0295)	(0.0281)	(0.0119)
weekend/holiday	0.1685***	0.1770***	0.1085***	0.0682***
	(0.0148)	(0.0150)	(0.0144)	(0.0069)
Constant	3.4715***	1.6703***	1.3677***	0.1500**
	(0.2103)	(0.1930)	(0.1248)	(0.0640)
var(ID)	-3.1803***	-3.0792***	-2.8808***	-3.4704***
	(0.1447)	(0.1346)	(0.1030)	(0.1054)
var(Residual)	-1.2345***	-1.2214***	-1.1516***	-1.8903***
	(0.0171)	(0.0156)	(0.0191)	(0.0231)
shift type effects	yes	yes	yes	yes
city of shift start	yes	yes	yes	yes
month effects	yes	yes	yes	yes
Observations	6147	6134	6128	6120

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

intervention and various time variables (i.e. month or duration of the experiment), indicating that the effect of real-time feedback is stable over time see Appendix D. Our results are also not a consequence of changes in the workforce. The analysis only includes employees who were recruited at least 1 month before the start of the experiment and we do not observe a drop in poor-performing employees or other changes across the treatments over the study period.

In contrast to messages containing social performance information, personal information alone had a positive, but not significant effect on sales productivity. Our first hypothesis is therefore only partially confirmed. However, the results are in line with Hypothesis 2, stating that the expected effects are lowest for the personal information group and strongest for the personal and social information treatment. We presume that this result is also driven by the fact that personal average performance levels are less informative or novel as stewards could also track this information themselves. Furthermore, the personal feedback information is not bonus-relevant. While we are not able to clearly separate the role of relative incentives in our setting, the complexity of the reward system and the fact that monetary rewards are not directly tied to the feedback content in any of the treatment groups suggest that the performance increases are rather intrinsically driven. Interestingly, we do not observe a more powerful impact of our treatments towards the end of the month, when social performance information (showing the last 30-day averages) is closest to the actual performance benchmarks used for bonus calculation.[13] We therefore assume that psychological factors such as self-satisfaction and self-efficacy (Bandura, 1988; Bandura & Cervone, 1983) or conformity (Bernheim, 1994) are better suited for explaining our results than the prospect of monetary rewards.

[13]The regression coefficients for the treatment and day of month-interactions lie between -0.036% (personal info) and 0.059% (personal and social info) with a p-value between 0.69 and 0.78. The Wald-tests are consequently not significant.

5.3 Real-time feedback effects and ability

As mentioned in Section 4, we presume different reactions to the feedback messages, depending on the stewards' general level of performance. To test this hypothesis, we split the minibar stewards in four quartiles, based on their sales performance before the study period: worst 25%, worse 25%, better 25% and best 25% (see 3.1 for calculation details). By using the stewards' past sales performance, instead of a more recent or dynamic performance measure as a basis, we are able to uniquely assign each employee to one of the four quartiles and avoid endogenous interactions with our intervention. With reference to model (1), we estimate the following interaction effect:

$$Ln(revhour)_{ijt} = \beta_0 + \beta_1 Group_i * Quartile_i + \beta \boldsymbol{Stew'}_i + \beta \boldsymbol{Shift'}_j + \beta \boldsymbol{Date'}_t + \beta \boldsymbol{Service'}_{ijt} + \upsilon_{0i} + \epsilon_{0ijt} \quad (2)$$

Table 3 shows the mixed effects estimation of model (2). In line with hypothesis 3, the interaction effects are particularly high and significant for those quartiles around the median. The personal and social feedback message in the second quartile (worse 25%), for example, leds to a revenue increase of almost 14%, compared to the worst 25% in the control group. As the first three rows of the regression output show, the treatment effects for the poorest performers tend to be negative. The coefficients for the best 25% are positive, but for the most part not significant.

For a further evaluation of the treatment effect for the different performance quartiles, Figure 3 shows the differences in the predicted margins of model (2). Differences in margins show the marginal effect of a certain value of the treatment variable compared to the control group, keeping everything else constant. Group differences in the log revenue per hour are particularly high and significant for those stewards who usually perform just below the average. Here, we observe an productivity increase of up to 10% compared to the control

Table 3: Performance Quartiles Regression

	log revenue/hour
personal info	-0.0302
	(0.0329)
social info	-0.0199
	(0.0227)
personal + social	-0.0405
	(0.0267)
worse 25% x personal info	0.0998**
	(0.0450)
worse 25% x social info	0.0965**
	(0.0401)
worse 25% x personal + social	0.1391***
	(0.0469)
better 25% x personal info	0.0598
	(0.0497)
better 25% x social info	0.0830***
	(0.0321)
better 25% x personal + social	0.1153***
	(0.0393)
best 25% x personal info	0.0447
	(0.0447)
best 25% x social info	0.0333
	(0.0313)
best 25% x personal + social	0.0679*
	(0.0351)
var(ID)	-3.3184***
	(0.1568)
var(Residual)	-1.2325***
	(0.0167)
steward fixed effects	yes
shift fixed effects	yes
date fixed effects	yes
service fixed effects	yes
Observations	6149

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

group. On the other hand, all treatments seem to have no or an even negative influence on sales for the best and poorest-performing stewards. Similar results appear when using the difference-in-difference approach as discussed in Section 5.2 (see Appendix E).

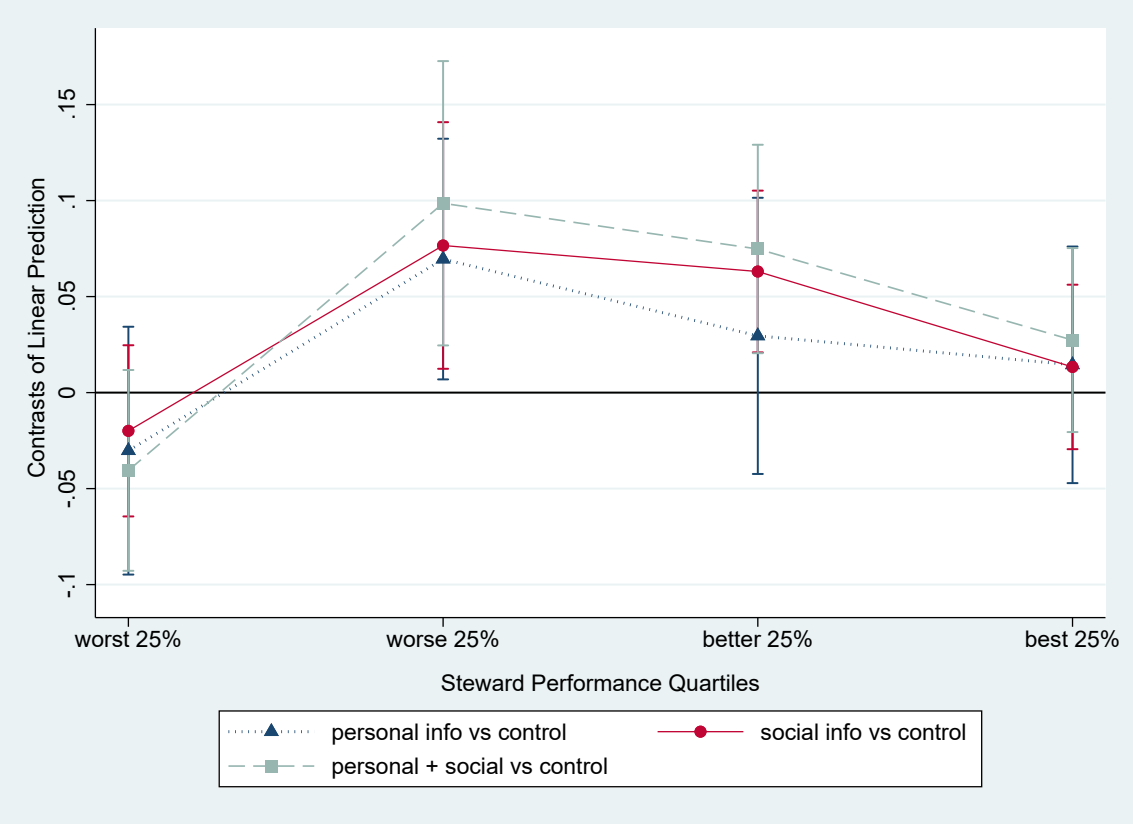

Figure 3: Contrasts of predictive margins with 95% Cis

While these results basically meet our expectations as stated in Hypothesis 3, some outputs stimulate further discussion. Interestingly, we observe a similar pattern for all feedback messages in each performance quartile. While this finding is initially rather surprising, it supports our previous point that sales increases are not only driven by potential bonus payments, but also by psychological and ability-related factors. In particular, the personal information treatment, which does not offer any reward-related performance information, also has a positive impact for the second performance quartile and a rather negative effect on stewards with very low abilities. Independently of the message content, the frequent tracking of revenues and the enhanced attention paid towards performance seems to moti-

vate below-average performers to realize their potential, while it has a presumably daunting effect on the poorest employees. The negative effect of all feedback message in the first quartile is also supported by the company's statement that direct superiors exert considerable performance pressure, particularly on poorly performing stewards.

On the other side of the performance distribution, none of the feedback messages results in a revenue increase for the top 25%-performers. Although the monthly bonus is divided proportionally to the stewards relative performance and additional effort would therefore pay off, we observe no significant effects in this quartile. In line with previous evidence (see Eriksson et al., 2009), we explain this result by a ceiling effect, suggesting that the top 25% stewards have already been working close to their performance limit.

6 Discussion

Practitioners are increasingly recognizing the benefits of setting small-scale goals and continuous performance evaluations (see Duggan, 2015). Yet, except some studies in the context of distribution centres (e.g. Goomas, 2012; T. D. Ludwig & Goomas, 2009), scientific evidence around real-time feedback is still scarce. The present study is one of the first contributions in the field, confirming that the automated provision of frequent and immediate performance information can lead to a significant productivity increase, beyond what is achieved by coarse, ex-post performance reviews. In the presence of a relative incentive scheme, our results demonstrate a growth in sales productivity of up to 3.8% when employees are regularly informed about personal and co-worker-related performance averages on their current work task (in addition to a monthly performance summary). Co-worker-related performance information alone leads to a similar productivity increase. These results are in line with existing evidence around social comparative information and rank feedback, which suggests that giving people the opportunity to compare themselves to others can lead to consider-

able productivity gains (Blanes i Vidal & Nossol, 2011; Delfgaauw et al., 2013; Kuhnen & Tymula, 2012; Schultz, 1999; Schultz et al., 2007; Szymanski & Harkins, 1987; White et al., 1995). Furthermore, these findings also support prior work demonstrating the benefits of more frequent or immediate feedback practices (Alavosius & Sulzer-Azaroff, 1986; Goomas et al., 2011; Kang et al., 2005; Mason & Redmon, 2008; Northcraft et al., 2011; So et al., 2013). Providing real-time feedback about personal performance standards solely, on the other hand, has no significant effect on performance. This type of information is commonly less novel and, in our setting, not bonus-relevant.

From a practical perspective, the monetary gains from real-time performance feedback in comparison to vague performance appraisals are quite substantial: An increase of 3.8% in revenues per hour in our study equals approximately 34'000 CHF additional revenue per month. Interestingly, and important from a practical point of view, the positive impact of co-worker-related real-time feedback does not seem to fade out over time. Since the effect appears stable, the monthly gains correspond to a revenue growth of more than 400'000 Swiss Francs per year. Considering that the one-time expenses for our intervention were 15'000 CHF for message programming, this productivity gain comes at almost no costs. As the existing incentive scheme is based on relative performance, the company does also not face additional bonus expenses.

The productivity growth can be mainly traced back to the fact that employees sell more products to a larger number of customers, rather than selling more expensive articles or intensifying cross-selling-activities. This finding is supported by earlier studies, revealing that competitive incentives may induce agents to work harder, but not necessarily smarter (Bracha & Fershtman, 2013). In a sales context, Casas-Arce and Martínez-Jerez (2009) demonstrate that retailers respond to tournaments by channelling most of the increased effort towards reaching more customers, instead of choosing smarter sales approaches. Our study indicates that this behavioural pattern also holds for the mere adaption of feedback

practices, without providing additional monetary rewards.

The paper also offers insights into how real-time feedback interacts with employee's general level of performance. In line with our prediction, we could show that the productivity increase is driven by workers in the middle of the performance distribution, particularly by those who usually perform just below the median. On the other hand, we find no or even a slight negative effect among the top- and bottom performers. This finding corresponds to existing literature on "dynamic incentive effects" and self-efficacy theory, suggesting a non-linear relationship between performance-standard discrepancies and subsequent effort in settings with and without relative rewards (Bandiera et al., 2013; Bartel, 2004; Casas-Arce & Martínez-Jerez, 2009; Delfgaauw et al., 2014; Eriksson et al., 2009; Kuhnen & Tymula, 2012). Due to these heterogeneous effects, firms might consider to strategically target real-time feedback. Following the approaches of prior studies, feedback could for example be provided more frequently to workers who possess average ability, offered to a subset of employees or the reference group of the feedback content might be adapted (see Kuhnen & Tymula, 2012).

With reference to the "dynamic incentive effect", previous studies have also demonstrated that the underlying reward system can have a strong influence on employees' reaction to feedback practices (see Hannan et al., 2008). Therefore, two specific characteristics of our setting should be considered when interpreting the present results: First, in contrast to the majority of previous studies, the relative incentives in our setting do rather resemble a multi-stage proportional prize contest (see Cason et al., 2010) than a tournament with one or a few winners. This feature has presumably mitigated a negative effect among workers at the top of the performance distribution. Secondly, the link between performance feedback and monetary rewards in our setting is rather weak, meaning that participants cannot directly infer clear monetary consequences from behavioural changes. This can be considered as another supportive factor in our design, since prior empirical and theoretical contributions

have proposed partial disclosures policies or coarse feedback, in order to maintain incentives for participants at the extremes of performance (e.g. Goltsman & Mukherjee, 2011; Hannan et al., 2008).

While relative incentives probably play some role in explaining our results, we assume a strong influence of psychological and ability-related factors. Existing evidence has shown that self-confidence, self-esteem and self-efficacy, constituting important drivers of motivation, are often determined by relative performance comparisons (see Benabou & Tirole, 2002; Köszegi, 2006; Szymanski & Harkins, 1987). Disentangling the role of different incentive structures and other, non-monetary factors with respect to the effects of real-time feedback is one of the key issues to be addressed in future research.

Looking ahead, we aim to further investigate the data at hand to also draw conclusions about employees' reaction to feedback messages over daytime and the immediate influence of benchmark achievements on work performance. One goal of further data analyses is also to determine the impact of different performance-standard discrepancies during work. As firms are increasingly adapting their feedback practices to exploit additional revenues, these and many other questions related to real-time feedback remain of great practical and academic interest.

7 Addition: Immediate performance effects of real-time feedback

This section provides an in-depth analysis of employees' immediate reaction to during-work feedback (see 3.2). By exploiting available data on single sales, we aim to investigate how (relative) performance information affects immediate work performance, directly after its release. These analyses should offer additional insights regarding the optimal timing of real-time performance feedback.

7.1 Introduction

Although regular performance feedback is a major trend in the business and private domain (e.g. fitness trackers and health gadgets), the *immediate* effects of real-time feedback on performance remain largely unknown. Several researchers empirically investigated the role of feedback frequency and immediacy in general (see 2.3. This literature shows an overall positive effect of more immediate feedback on performance, but does not address the direct impact of during-work feedback over time. In a field experiment, Houde, Todd, Sudarshan, and Carrie Armel (2013) study the effect of real-time feedback on electricity consumption over daytime. However, they focus on day-related consumption patterns, without considering the time of feedback release and its immediate impact.

Partially related to our analyses is the growing literature on interim performance feedback. Lizzeri, Meyer, and Persico (2002) study a principal-agent model with two periods, in which the agent cannot fully observe his performance. Under a predefined incentive scheme, they show that the agent's total expected effort is higher in some circumstances when his first-period outcome is revealed. In a two-stage tournament, S. Ludwig and Luenser (2008) find that intermediate feedback does not influence subjects' second stage effort choices by

itself, but conditional in their relative performance. Those participants who lag tend to increase their second stage effort and those who lead tend to decrease it. Intermediate information therefore leads to a more balanced tournament. In a a similar setting, Aoyagi (2010) and Ederer (2010) suggest that the optimal disclosure policy depends on the agent's cost of effort function. If the marginal cost of effort in stage two is convex, the no-interim feedback policy is optimal. On the other hand, the full information revelation is optimal when the stage two marginal cost is concave. Based on the assumption that agents know about their ability and this knowledge enters the production function, Ederer (2010) further distinguishes between the beneficial "motivation effect" and the adverse "evaluation effect" of interim feedback. While interim information helps the agent in tailoring effort to his correct ability level, it also reveals how likely an agent is to win the tournament, which is likely to have a negative impact in case of a large performance gap. Firms therefore face a fundamental trade-off when deciding whether to provide interim performance information. Goltsman and Mukherjee (2011) show that feedback disclosure policies that enhance final-stage effort may dampen incentives at the intermediate stage. They model an optimal feedback policy where the firm discloses information only if both workers fail at the intermediate stage and does not disclose any information following any other intermediate outcome.

With a closer link to our question, Eriksson et al. (2009) explore the effect of peer-related performance information during a calculation task. Given a tournament incentive scheme, they show that intermediate information regarding the competitor's performance has no effect on a participant's task fulfilment in the second stage. Only the continuous revelation of peer-feedback leads to a performance increase of the losing player in the second half of the task. This motivational effect can be observed if the score gap is not too high. Leading players, on the other hand, are not affected in neither feedback condition and do not adjust their performance to the distance with the underdog.

Like Eriksson et al. (2009), we analyse the effect of intermediate feedback, but in our

experiment the provided information is not novel. Stewards get to know the respective performance benchmarks at the beginning of their shift and they can access their current revenue at any time during work via the electronic till. We therefore investigate the immediate effect of interim feedback, which is made more salient. There exists some prior evidence that the salience of already known information has an effect on work performance (see Englmaier, Roider, & Sunde, 2017).

Our hypothesis is that salient during-work feedback leads to a performance increase directly after its release. Based on the literature around feedback and ability (see 2.2) and interim information revelation, we further expect the temporal effect of the feedback message to depend on an employee's present performance level. We assume a positive, immediate impact on performance when the benchmark of the feedback message is perceived as difficult but attainable. On the other hand, we expect a negative, immediate performance impact if it becomes visible that an employee has already achieved the benchmark or is highly likely to achieve it by the end of the shift.

7.2 Empirical strategy

The during-work feedback message, which was originally programmed to appear at a random time during the service, was not trackable. These messages were therefore reprogrammed on April 21 2016, seven weeks after the study start. For the remaining ten weeks of the experiment, the during-service messages were released according to a pre-defined timetable: two hours after shift start (steward login) in the first week, three hours after login in second week, four hours in the third week, then again after two hours and so forth. Due to differing starting times of the shifts, the during-work messages on a certain date appeared at different daytimes. Figure 4 provides an illustration.

Based on this design, we are able to make a within-treatment comparison between those

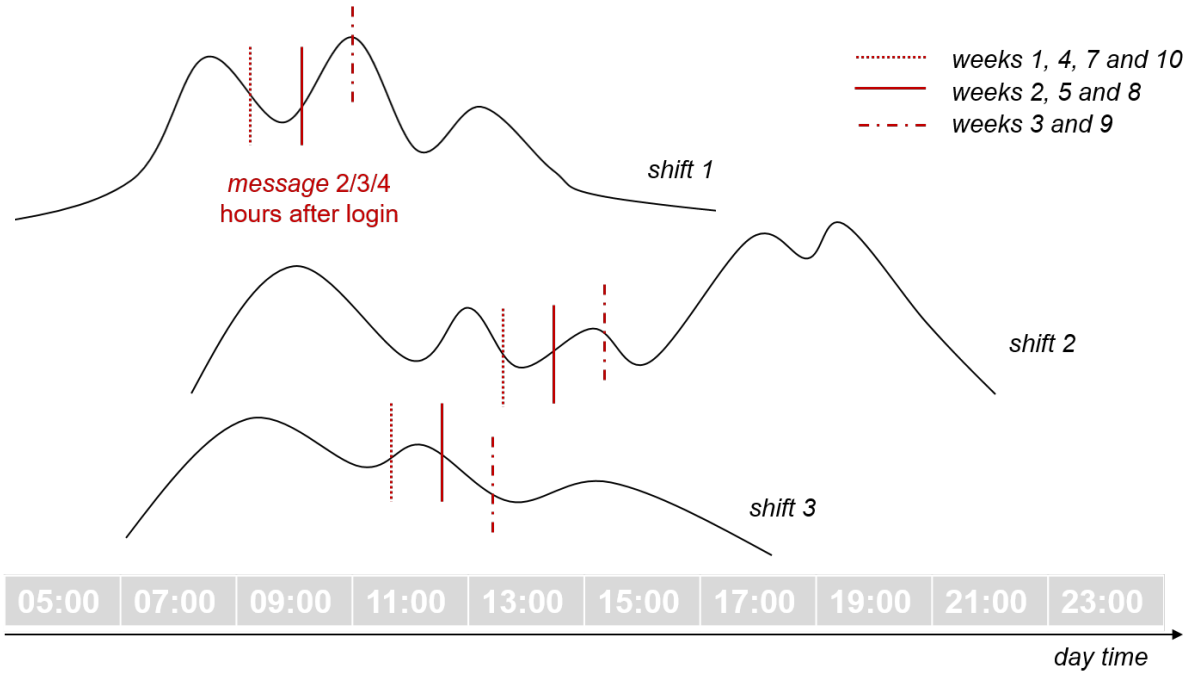

Figure 4: Illustration of the during-work feedback

stewards who have received the during-service feedback and those stewards, who did not (yet) obtain the message at a certain point of daytime. More specifically, we compare the sales performance of observations, where the same steward on the same shift did receive the feedback message during the past 60 minutes and where he/she did not yet receive the during-work feedback. Figure 5 illustrates the treatment (green) and control groups (red). Since we are missing a clean control group for the "message after 4 hour"-condition, we confine our analysis to the third and fourth working hour after login.

7.3 Data and model

In order to investigate time-related effects, we use data on the level of every single sale. For each service (i.e. a shift performed by a certain steward, on a certain date), we split up the single sales into multiple time frames of one hour. The first time frame starts at the login time of the steward. Since we have 0 sales for around 13% of these observations, we use the number of articles sold per hour, instead of the logarithmised sales revenue as our

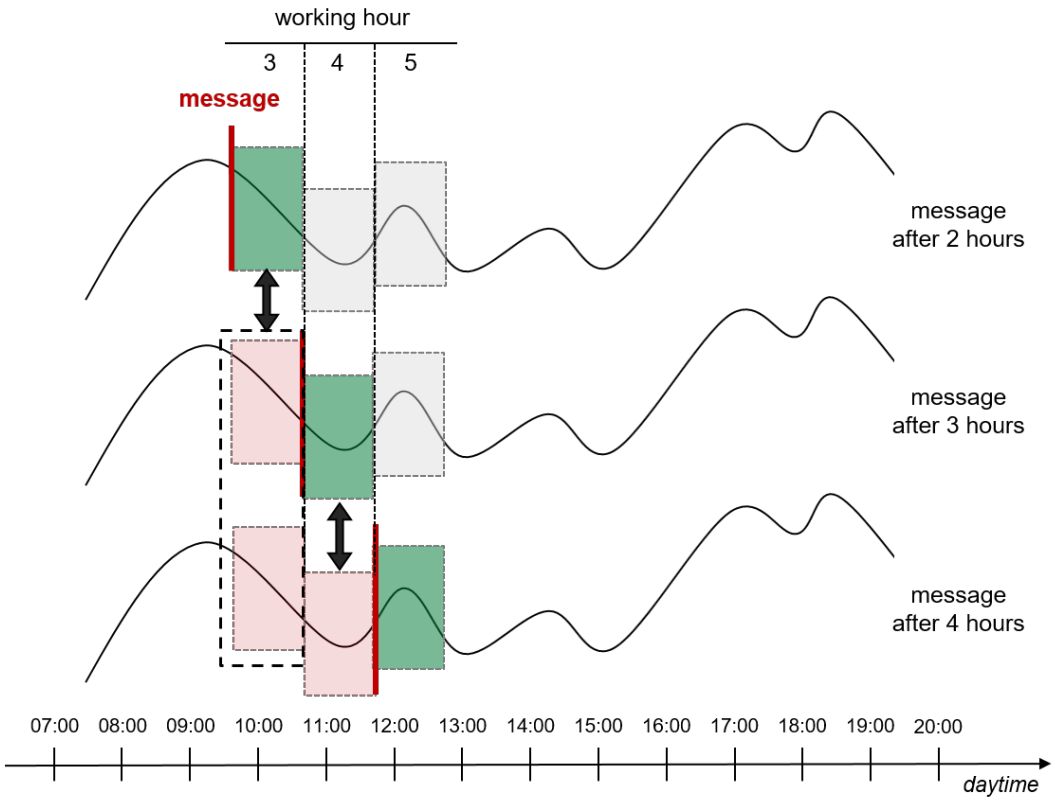

Figure 5: Model illustration

main outcome variable. The number of sales by steward i on service j in working hour t is estimated with the following poisson regression model. Further alternatives, such as a multi-level model and a negative binomial regression are provided in Appendix **??**.

$$Log(E(Sales)_{ijt}) = \beta_0 + \beta_1 \mathbf{message}_{ijt} + \beta steward'_i + \beta service'_j +$$
$$\beta_2 daytime_t + \beta_3 workingtime_{ijt} + \beta_4 occupancy_{jt} + \beta_5 break\ time_{jt} +$$
$$\beta dayofweek'_j + \beta month'_j + \epsilon_{0ijt} \quad (3)$$

message is a dummy indicator for the message release at the beginning of working hour t. Beside steward- and shift-fixed effects, we also control for daytime, the duration for which a stewards has already been working (working hour), the average occupancy rate of the train during hour t (in %) and break time during t (in minutes). We further include some date-

related variables, which possibly influence sales performance, such as the day of the week and month.

Our analysis includes all minibar services of the treatment groups (see 3.2) from April 21 until June 30, 2016. Observations were excluded if the during-service message appeared during a break, during a change of trains or if there was a train failure at any time during the service. As mentioned before, we further confine our analysis to the third and fourth working hour, where some stewards did already receive the during-service feedback and others did not (see Figure 4). This leads to a total of 2150 working hour observations (1075 services with two working hours each). In 40% of these cases, the during-service message appeared two hours after login. The relative amount of observations for the three and four hours after login-conditions are 28% and 32%.

7.4 Results

Table 4 shows the estimates for different specifications of model 3. Taking all treatment groups together, we find a slightly negative, but not significant effect of the during-service message on the number of articles sold in the subsequent hour. As Appendix F shows, the results stay almost the same if we spit up the data into half-hour time frames and look at sales performance within 30 (instead of 60) minutes after feedback release.

The coefficients for break time and occupancy rate both point into the expected direction. The positive impact of the fourth, in comparison to the third working hour, may be explained by a clear peak in break time in hour three. Stewards then possibly have additional energy or motivation in hour four. The right part of the table shows separate estimates of the full specification (3) for the single treatment groups. These results reveal a significant negative effect of the social info feedback on immediate sales performance. Within this treatment, the expected number of sales is $(e^{0.169} - 1) * 100 = 15.5\%$ lower, if a steward has received the

during-work message at the beginning of hour t. Messages containing a personal performance benchmark (treatment 1 and 3) do not seem to have an immediate impact. Appendix G provides two separate estimates for working hour three and four. Overall, the immediate performance effect is not significant in neither of the time frames, but a bit more negative for working hour four (when the message was shown after three hours).

For a further understanding of these effects, we also analyse the influence of a steward's *current* performance at the time of feedback release [14]. We measure this present performance level by the remaining revenue per residual working hour that a steward has to generate in order to achieve the feedback benchmark. In terms of the personal (social) info treatment, this benchmark constitutes the personal (total), average revenue of the (all) steward(s) on the same shift during the past 30 days. For the "personal+social" treatment, we chose the social benchmark to calculate the current performance measure.[15] By taking into account the remaining working time, we ensure that the performance measure is independent from the time of feedback disclosure (two, three or four hours after login).

[14] While we were looking at the interaction with a steward's general ability in Section 5.3, the focus here is on a worker's day-related performance.

[15] As previous analyses revealed, personal performance information has a far lower impact on performance (see Section 5.2).

Table 4: Poisson Regression: Articles sold per working hour

	All treatment groups			Single treatments		
	(1)	(2)	(3)	Personal Info	Social Info	Persona + Social
feedback message	-0.071*	-0.022	-0.032	0.057	-0.169***	-0.019
	(0.040)	(0.035)	(0.035)	(0.071)	(0.060)	(0.057)
working hour=4	0.114***	0.182***	0.270***	0.255*	0.227*	0.249*
	(0.041)	(0.032)	(0.066)	(0.131)	(0.121)	(0.151)
break (in min)	-0.032***	-0.035***	-0.034***	-0.030***	-0.037***	-0.035***
	(0.002)	(0.002)	(0.002)	(0.004)	(0.004)	(0.003)
occupancy (in %)	0.001	0.006***	0.008***	0.011***	0.006***	0.007***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Steward FE	No	Yes	Yes	Yes	Yes	Yes
Shift FE	No	Yes	Yes	Yes	Yes	Yes
Daytime FE	No	No	Yes	Yes	Yes	Yes
Month FE	No	No	Yes	Yes	Yes	Yes
Day of week FE	No	No	Yes	Yes	Yes	Yes
Pseudo R-sq	0.128	0.355	0.391	0.444	0.396	0.398
Observations	1994	1994	1994	581	755	658

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5 shows the estimates of model 3, including the interaction with a steward's current performance. Taking all treatment groups together, we can observe a significant negative impact of the during-work message at the time of benchmark achievement (when *rev to go is 0*). On the other hand, the during-work feedback effect points into a positive direction as the deviation from the feedback benchmark increases. This interaction effect is confirmed and even stronger in various, alternative model specifications. Appendix **??** shows the results when using a negative binomial regression and a mixed poisson model with random effects for each service. Both are common approaches for handling longitudinal (count) data with zeros (see Atkins, Baldwin, Zheng, Gallop, & Neighbors, 2013; Donald & Robert D., 2006). [16]

As the single-treatment results suggest, the feedback effect is particularly sensitive to a steward's current performance in the social info treatment. Here, the estimated decrease in subsequent sales is $(e^{0.434} - 1) * 100 = 32.2\%$ after an employee reached the performance benchmark. Congruent with previous findings of this paper, stewards in the personal info condition do not show any response to the during-work feedback message, independently of their present revenue. Figure 6 provides the predictive margins for the number of sales after the during-work message, in contrast to the predictive margins without the message for each treatment group. The immediate feedback effect in the "personal+social"-treatment becomes positive for stewards who have to earn more than 36.2 CHF per hour for the remaining working day. This is close to the median revenue to go per hour of 39.5 CHF. Within the social info treatment, the respective turning point lies at 82.8 CHF/h, implying that the immediate performance impact is positive in only 1-2% of the cases. Separate graphs per treatment including the confidence intervals are provided in Appendix H.

[16] We obtain a similar, linear interaction effect if we use current performance quartiles instead of *rev to go* as a continuous variable. The coefficients are, however, only marginally significant. Including the quadratic term *rev to go*^{2} in addition to *rev to go* does not improve the fit of the model. In contrast to the literature on dynamic incentive effects (see Section 2.2), we can therefore assume that employees do not slack off, even if their present performance is very poor.

Table 5: Poisson Regression: Articles sold per working hour

	All treatment groups			Single treatments		
	(1)	(2)	(3)	Personal Info	Social Info	Personal + Social
feedback message	-0.201**	-0.216***	-0.166**	0.006	-0.434***	-0.193*
	(0.081)	(0.073)	(0.068)	(0.134)	(0.126)	(0.105)
rev to go (CHF per h)	-0.003*	-0.011***	-0.009***	-0.009***	-0.014***	-0.007***
	(0.002)	(0.001)	(0.001)	(0.003)	(0.002)	(0.002)
feedback message x rev to go (CHF per h)	0.003*	0.005***	0.003**	0.001	0.005**	0.005**
	(0.002)	(0.002)	(0.001)	(0.003)	(0.003)	(0.002)
working hour=4	0.112***	0.136***	0.199***	0.239*	0.096	0.182
	(0.042)	(0.032)	(0.065)	(0.131)	(0.119)	(0.147)
break (in min)	-0.032***	-0.035***	-0.034***	-0.030***	-0.037***	-0.035***
	(0.002)	(0.002)	(0.002)	(0.004)	(0.003)	(0.003)
occupancy (in %)	0.000	0.005***	0.006***	0.009***	0.004*	0.006***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Steward FE	No	Yes	Yes	Yes	Yes	Yes
Shift FE	No	Yes	Yes	Yes	Yes	Yes
Daytime FE	No	No	Yes	Yes	Yes	Yes
Month FE	No	No	Yes	Yes	Yes	Yes
Day of week FE	No	No	Yes	Yes	Yes	Yes
Pseudo R-sq	0.130	0.369	0.400	0.454	0.413	0.402
Observations	1994	1994	1994	581	755	658

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

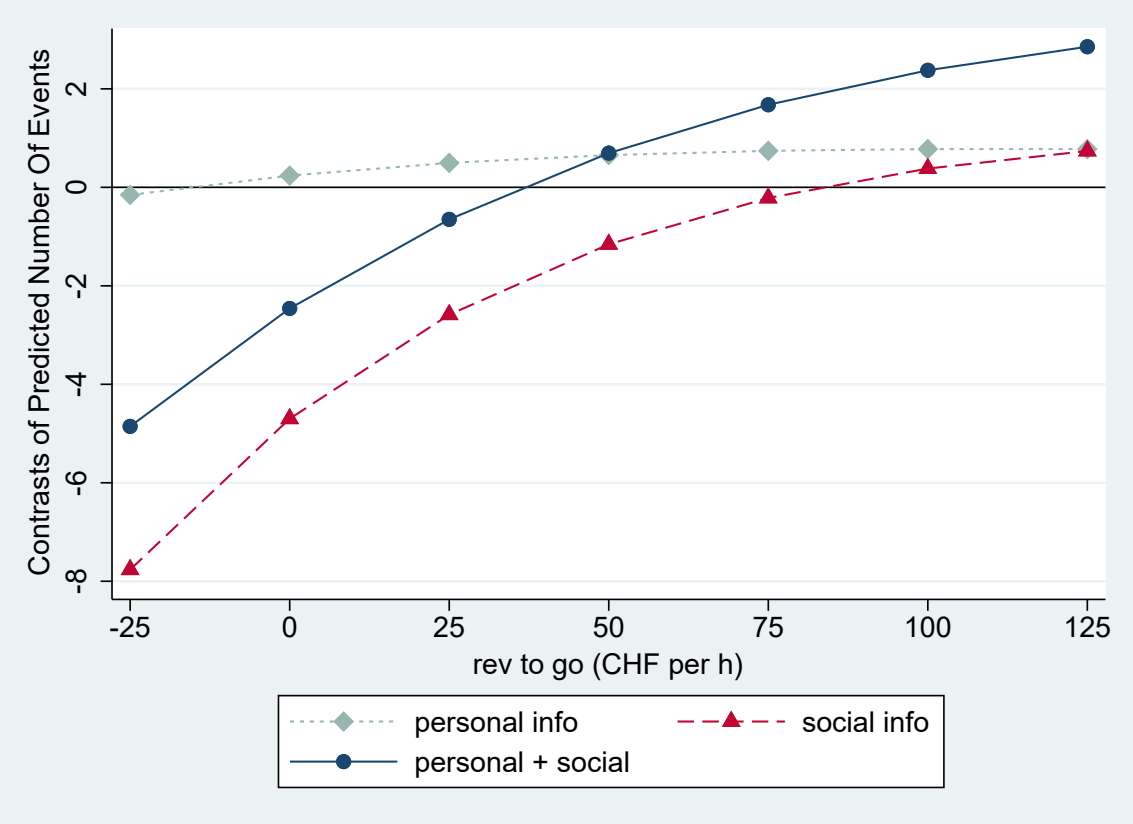

Figure 6: Contrasts of predictive margins for during work-feedback

Table 6 shows the performance-feedback interaction for working hour three and four separately. Interestingly, the interaction effect seems to be driven by working hour four, i.e. for feedback messages which were released later on in the shift.[17] Although we have are not able to track this trend further, this result indicates that the immediate effect is more sensitive to a worker's current performance if feedback is displayed towards the end of the shift. Earlier on the working day, the gap to the performance-benchmark seems to be less influential. This might be explained by the fact that early feedback is less informative or definite and subsequent performance may therefore be less (directly) adjusted than for late feedback.

[17] Graphs of the predictive margins per working hour can be found in Appendix H.

Table 6: Poisson Regression: Articles sold per working hour

	Working hour 3	Working hour 4
feedback message	-0.116	-0.357***
	(0.083)	(0.099)
rev to go (CHF per h)	-0.012***	-0.009***
	(0.002)	(0.002)
feedback message x rev to go (CHF per h)	0.002	0.006***
	(0.002)	(0.002)
break (in min)	-0.032***	-0.024***
	(0.005)	(0.007)
occupancy (in %)	0.006***	0.000
	(0.001)	(0.001)
Steward FE	Yes	Yes
Shift FE	Yes	Yes
Daytime FE	Yes	Yes
Month FE	Yes	Yes
Day of week FE	Yes	Yes
Pseudo R-sq	0.521	0.537
Observations	997	597

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

7.5 Discussion

While performance feedback is one of the most extensively studied fields in behavioral research, there is surprisingly little knowledge about how feedback affects performance immediately after its release and over the time of a task. Our analyses shed a first light on this question. Against our hypothesis, the above results do not confirm an immediate positive effect of during-work feedback on subsequent performance. Yet, employees react to salient performance information, directly after its release. For messages including social performance benchmarks, workers who perform far below the total average tend to be immediately motivated. Feedback which underlines that the performance benchmark is likely to be reached, however, has a significant negative impact on immediate performance. This is particularly the case, if well-performing workers do not receive a personal performance benchmark, they can additionally compete with, but social feedback solely. Overall, we find that the direct effectiveness of feedback containing social (or social and personal) performance information crucially depends on a worker's performance at the time of feedback release. Personal performance information alone does, however, not affect immediate sales, independently of an employee's current position.

This result is consistent with the literature around interim performance feedback, suggesting that an agent's reaction depends on his relative performance (e.g. Ederer, 2010; Goltsman & Mukherjee, 2011; S. Ludwig & Luenser, 2008). It also confirms our previous findings that employee performance is an important mediator for the effectiveness of real-time feedback and now provides a first indication that this interaction effect also holds for the immediate impact of feedback over daytime. The interaction even occurs in a setting where performance information is actually not novel, but only made more salient to employees. Furthermore, in contrast to the literature on intermediate information in tournaments, employees' performance is only indirectly tied to monetary rewards.

Our analyses also provide some preliminary evidence that the feedback-performance in-

teraction is not uniform. It rather appears that an employee's current performance has a stronger influence, if feedback is provided towards the end of the task. On the other hand, the immediate reaction seems to be less driven by the present attainment level an employee when feedback is provided at an early stage. We explain this results by the lower urgency of effort adjustments, when performance gaps become salient early on.

Although research around immediate feedback effects is still scarce, our findings provide some preliminary suggestions for practice. First, despite the overall positive effects of social performance information, making this type of feedback salient during work is no general mean for immediate, short-term productivity increases. Our findings rather suggest that social feedback during work should be provided selectively for current poor-performers, in order to prevent potential negative effects. Furthermore, if a selective provision of feedback is not feasible, it may be reasonable for companies to disclose vague performance information. Partial disclosure or coarse feedback, instead of full revelation of interim performance, has been proposed as an optimal strategy in previous studies (Goltsman & Mukherjee, 2011; Hannan et al., 2008). Companies may implement some kind of "partial disclosure policy" by providing feedback at an early stage of the task, where the final outcome is still indefinite and its information value is relatively low.

Looking ahead, the immediate effect of real-time feedback, its relation to current performance levels and the role of different release times require further research. Additional insights into these questions would be beneficial for an optimal timing of feedback messages in practice. This is particularly relevant, as information technologies provide ever wider options for customized feedback systems in both, the private and commercial spheres.

References

Akın, Z., & Karagözoğlu, E. (2017). The role of goals and feedback in incentivizing performance. *Managerial and Decision Economics*, *38*(2), 193–211.

Alavosius, M. P., & Sulzer-Azaroff, B. (1986). The effects of performance feedback on the safety of client lifting and transfer. *Journal of applied behavior analysis*, *19*(3), 261–267.

Alvero, A. M., Bucklin, B. R., & Austin, J. (2001). An objective review of the effectiveness and essential characteristics of performance feedback in organizational settings (1985-1998). *Journal of Organizational Behavior Management*, *21*(1), 3–29.

Aoyagi, M. (2010). Information feedback in a dynamic tournament. *Games and Economic Behavior*, *70*(2), 242–260.

Atkins, D. C., Baldwin, S. A., Zheng, C., Gallop, R. J., & Neighbors, C. (2013). A tutorial on count regression and zero-altered count models for longitudinal substance use data. *Psychology of addictive behaviors : journal of the Society of Psychologists in Addictive Behaviors*, *27*(1), 166–177.

Azmat, G., & Iriberri, N. (2010). The importance of relative performance feedback information. *Journal of Public Economics*, *94*(7-8), 435–452.

Azmat, G., & Iriberri, N. (2016). The provision of relative performance feedback. *Journal of Economics & Management Strategy*, *25*(1), 77–110.

Bandiera, O., Barankay, I., & Rasul, I. (2013). Team incentives: Evidence from a firm level experiment. *Journal of the European Economic Association*, *11*(5), 1079–1114.

Bandura, A. (1988). Self-regulation of motivation and action through goal systems. In V. Hamilton, G. H. Bower, & N. H. Frijda (Eds.), *Cognitive perspectives on emotion and motivation* (pp. 37–61). Dordrecht: Springer Netherlands.

Bandura, A., & Cervone, D. (1983). Self-evaluative and self-efficacy mechanisms governing the motivational effects of goal systems. *Journal of Personality and Social Psychology*, *45*(5), 1017–1028.

Bandura, A., & Cervone, D. (1986). Differential engagement of self-reactive influences in cognitive motivation. *Organizational Behavior and Human Decision Processes*, *38*(1), 92–113.

Bandura, A., & Jourden, F. J. (1991). Self-regulatory mechanisms governing the impact of social comparison on complex decision making. *Journal of Personality and Social Psychology*, *60*(6), 941–951.

Barankay, I. (2011). Rankings and social tournaments: Evidence from a crowd-sourcing experiment. *Wharton School of Business, University of Pennsylvania Working Paper*.

Bartel, A. P. (2004). Human resource management and organizational performance. *Industrial and Labor Relations Review*, *57*(2), 181.

Bartol, K. M., Durham, C. C., & Poon, J. M. L. (2001). Influence of performance evaluation rating segmentation on motivation and fairness perceptions. *Journal of Applied Psychology*, *86*(6), 1106–1119.

Benabou, R., & Tirole, J. (2002). Self-confidence and personal motivation. *The quarterly journal of economics*, *117*(3), 871–915.

Bernheim, B. D. (1994). A theory of conformity. *Journal of political economy*.

Blader, S., Gartenberg, C. M., & Prat, A. (2015). The contingent effect of management practices. *SSRN Electronic Journal*.

Blanes i Vidal, J., & Nossol, M. (2011). Tournaments without prizes: Evidence from personnel records. *Management Science*, *57*(10), 1721–1736.

Bracha, A., & Fershtman, C. (2013). Competitive incentives. *Management Science*, *59*(4), 771–781.

Braunstein, D. N., Klein, G. A., & Pachla, M. (1973). Feedback expectancy and shifts in student ratings of college faculty. *Journal of Applied Psychology*, *58*(2), 254–258.

Bullen, N., Jones, K., & Duncan, C. (1997). Modelling complexity. *Environment and Planning A*, *29*(4), 585–609.

Buunk, A. P., & Gibbons, F. X. (2007). Social comparison. *Organizational Behavior and Human Decision Processes*, *102*(1), 3–21.

Cameron, A. C., & Trivedi, P. K. (2010). *Microeconometrics using stata* (Rev. ed. ed.). College Station, Tex.: Stata Press.

Campion, M. A., & Lord, R. G. (1982). A control systems conceptualization of the goal-setting and changing process. *Organizational Behavior and Human Performance*, *30*(2), 265–287.

Cappelli, P., & Tavis, A. (2016). The performance management revolution. *Harvard Business Review*, *10*, 58–67. Retrieved from `https://hbr.org/2016/10/the-performance-management-revolution`

Casas-Arce, P., Lourenço, S. M., & Martínez-Jerez, F. A. (2017). The performance effect of feedback frequency and detail. *Journal of Accounting Research*, *62*, 151.

Casas-Arce, P., & Martínez-Jerez, F. A. (2009). Relative performance compensation, contests, and dynamic incentives. *Management Science*, *55*(8), 1306–1320.

Cason, T. N., Masters, W. A., & Sheremeta, R. M. (2010). Entry into winner-take-all and proportional-prize contests: An experimental study. *Journal of Public Economics*, *94*(9-10), 604–611.

Chhokar, J. S., & Wallin, J. A. (1984). A field study of the effect of feedback frequency on performance. *Journal of Applied Psychology*, *69*(3), 524–530.

Colin Cameron, A., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, *50*(2), 317–372.

Crowell, C. R., Anderson, D. C., Abel, D. M., & Sergio, J. P. (1988). Task clarification, performance feedback, and social praise. *Journal of Applied Behavior Analysis*, *21*(1), 65–71.

Delfgaauw, J., Dur, R., Non, A., & Verbeke, W. (2014). Dynamic incentive effects of relative performance pay. *Labour Economics*, *28*, 1–13.

Delfgaauw, J., Dur, R., Sol, J., & Verbeke, W. (2013). Tournament incentives in the field. *Journal of labor economics*, *31*(2), 305–326.

Deloitte University Press (Ed.). (2015). *Global human capital trends 2015.* Retrieved from `https://www2.deloitte.com/content/dam/Deloitte/de/Documents/human-capital/HCTrends%202015%20Report_TuesFeb24.pdf`

Donald, H., & Robert D., G. (Eds.). (2006). *Longitudinal data analysis.* Hoboken, NJ, USA: John Wiley & Sons, Inc.

Duggan, K. (2015). Why the annual performance review is going extinct. *Fast Company*. Retrieved from `https://www.fastcompany.com/3052135/why-the-annual-performance-review-is-going-extinct`

Earley, P. C., Northcraft, G. B., Lee, C., & Lituchy, T. R. (1990). Impact of process and outcome feedback on the relation of goal setting to task performance. *Academy of Management Journal*, *33*(1), 87–105.

Ederer, F. (2010). Feedback and motivation in dynamic tournaments. *Journal of economics & management strategy : JEMS*, *19*(3), 733–769.

Englmaier, F., Roider, A., & Sunde, U. (2017). The role of communication of performance schemes. *Management Science*, *63*(12), 4061–4080.

Eriksson, T., Poulsen, A., & Villeval, M. C. (2009). Feedback and incentives: Experimental evidence. *Labour Economics*, *16*(6), 679–688.

Fajfar, P., Campitelli, G., & Labollita, M. (2012). Effects of immediacy of feedback on estimations and performance. *Australian Journal of Psychology*, *64*(3), 169–177.

Feather, N. T. (Ed.). (1982). *Expectations and actions: Expectaney-value models in psychology.* Hillsdale, New Jersey: Lawrence Elbaum Associates.

Fehr, E., & Goette, L. (2007). Do workers work more if wages are high? *American Economic Review*, *97*(1), 298–317.

Fischer, C. (2008). Feedback on household electricity consumption. *Energy Efficiency*, *1*(1), 79–104.

Gino, F., & Staats, B. R. (2011). Driven by social comparisons. *SSRN Electronic Journal*.

Goltsman, M., & Mukherjee, A. (2011). Interim performance feedback in multistage tournaments. *Journal of labor economics*, *29*(2), 229–265.

Goomas, D. T. (2012). Immediate feedback on accuracy and performance. *Journal of Organizational Behavior Management*, *32*(4), 320–328.

Goomas, D. T., & Ludwig, T. D. (2007). Enhancing incentive programs with proximal goals and immediate feedback. *Journal of Organizational Behavior Management*, *27*(1), 33–68.

Goomas, D. T., Smith, S. M., & Ludwig, T. D. (2011). Business activity monitoring. *Journal of Organizational Behavior Management*, *31*(3), 196–209.

Hannan, R. L., Krishnan, R., & Newman, A. H. (2008). The effects of disseminating relative performance feedback in tournament and individual performance compensation plans. *The accounting review : a journal of the American Accounting Association*, *83*(4), 893–913.

Heckhausen, H. (1977). Achievement motivation and its constructs. *Motivation and Emotion*, *1*(4), 283–329.

Houde, S., Todd, A., Sudarshan, A., & Carrie Armel, K. (2013). Real-time feedback and electricity consumption. *The Energy Journal*, *34*(1).

Hundal, P. S. (1969). Knowledge of performance as an incentive in repetitive industrial work. *Journal of Applied Psychology*, *53*(3, Pt.1), 224–226.

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*, *64*(4), 349–371.

Kang, K., Oah, S., & Dickinson, A. M. (2005). The relative effects of different frequencies of feedback on work performance. *Journal of Organizational Behavior Management*, *23*(4), 21–53.

Kettle, K. L., & Häubl, G. (2010). Motivation by anticipation: expecting rapid feedback enhances performance. *Psychological science*, *21*(4), 545–547.

Kim, J. S., & Hamner, W. C. (1976). Effect of performance feedback and goal setting on productivity and satisfaction in an organizational setting. *Journal of Applied Psychology*, *61*(1), 48–57.

Klein, W. M. (1997). Objective standards are not enough. *Journal of Personality and Social Psychology*, *72*(4), 763–774.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254–284.

Köszegi, B. (2006). Ego utility, overconfidence, and task choice. *Journal of the European*

Economic Association, *4*(4), 673–707.

Kuhnen, C. M., & Tymula, A. (2012). Feedback, self-esteem, and performance in organizations. *Management Science*, *58*(1), 94–113.

Latham, G. P., & Locke, E. A. (1991). Self-regulation through goal setting. *Organizational Behavior and Human Decision Processes*, *50*(2), 212–247.

Lizzeri, A., Meyer, M. A., & Persico, N. (2002). *The incentive effects of interim performance evaluations* (Vols. 02,09).

Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal setting and task performance: 1969-1980. *Psychological Bulletin*, *90*(1), 125–152.

Ludwig, S., & Luenser, G. (2008). *Knowing the gap-intermediate information in tournaments*. Unpublished Manuskript.

Ludwig, T. D., & Goomas, D. T. (2009). Real-time performance monitoring, goal-setting, and feedback for forklift drivers in a distribution centre. *Journal of Occupational and Organizational Psychology*, *82*(2), 391–403.

Lurie, N. H., & Swaminathan, J. M. (2009). Is timely information always better? *Organizational Behavior and Human Decision Processes*, *108*(2), 315–329.

Mason, M. A., & Redmon, W. K. (2008). Effects of immediate versus delayed feedback on error detection accuracy in a quality control simulation. *Journal of Organizational Behavior Management*, *13*(1), 49–83.

Moore, D. A., & Klein, W. M. (2008). Use of absolute and comparative performance feedback in absolute and comparative judgments and decisions. *Organizational Behavior and Human Decision Processes*, *107*(1), 60–74.

Nolan, T. V., Jarema, K. A., & Austin, J. (1999). An objective review of the journal of organizational behavior management. *Journal of Organizational Behavior Management*, *19*(3), 83–114.

Nordstrom, R., Hall, R. V., Lorenzi, P., & Delquadri, J. (1988). Organizational behavior modification in the public sector. *Journal of Organizational Behavior Management*, *9*(2), 91–112.

Northcraft, G. B., Schmidt, A. M., & Ashford, S. J. (2011). Feedback and the rationing of time and effort among competing tasks. *The Journal of applied psychology*, *96*(5), 1076–1086.

Schultz, P. W. (1999). Changing behavior with normative feedback interventions. *Basic and Applied Social Psychology*, *21*(1), 25–36.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological science*, *18*(5), 429–434.

Schunk, D. H., & Swartz, C. W. (1993). Goals and progress feedback. *Contemporary Educational Psychology*, *18*(3), 337–354.

Sharma, D. A., Chevidikunnan, M. F., Khan, F. R., & Gaowgzeh, R. A. (2016). Effectiveness of knowledge of result and knowledge of performance in the learning of a skilled motor activity by healthy young adults. *Journal of physical therapy science*, *28*(5), 1482–1486.

So, Y., Lee, K., & Oah, S. (2013). Relative effects of daily feedback and weekly feedback on customer service behavior at a gas station. *Journal of Organizational Behavior Management*, *33*(2), 137–151.

Son, H. (2017-03-09). At jpmorgan, your performance review is now. and now. and now... *Bloomberg*. Retrieved from `https://www.bloomberg.com/news/articles/2017-03-09/at-jpmorgan-your-performance-review-is-now-and-now-and-now`

Strang, H. R., Lawrence, E. C., & Fowler, P. C. (1978). Effects of assigned goal level and knowledge of results on arithmetic computation. *Journal of Applied Psychology*, *63*(4), 446–450.

Subramanian, S. V., Jones, K., Kaddour, A., & Krieger, N. (2009). Revisiting robinson: the perils of individualistic and ecologic fallacy. *International journal of epidemiology*, *38*(2), 342-60; author reply 370-3.

Surane, J. (2017-04-21). Goldman sachs introduces real-time employee performance reviews. *Bloomberg*. Retrieved from `https://www.bloomberg.com/news/articles/2017-04-21/goldman-sachs-introduces-real-time-employee-performance-reviews`

Szymanski, K., & Harkins, S. G. (1987). Social loafing and self-evaluation with a social standard. *Journal of Personality and Social Psychology*, *53*(5), 891–897.

The Economist. (2016). *The measure of a man.* Retrieved from `https://www.economist.com/news/21693151-employers-are-modifying-not-abolishing-them-performance-reviews-not-dead-yet`

Tosi, H. L., Locke, E. A., & Latham, G. P. (1991). A theory of goal setting and task performance. *The Academy of Management Review*, *16*(2), 480.

Tran, A., & Zeckhauser, R. (2012). Rank as an inherent incentive. *Journal of Public Economics*, *96*(9-10), 645–650.

White, P. H., Kjelgaard, M. M., & Harkins, S. G. (1995). Testing the contribution of self-evaluation to goal-setting effects. *Journal of Personality and Social Psychology*, *69*(1), 69–79.

Appendix A Employee performance

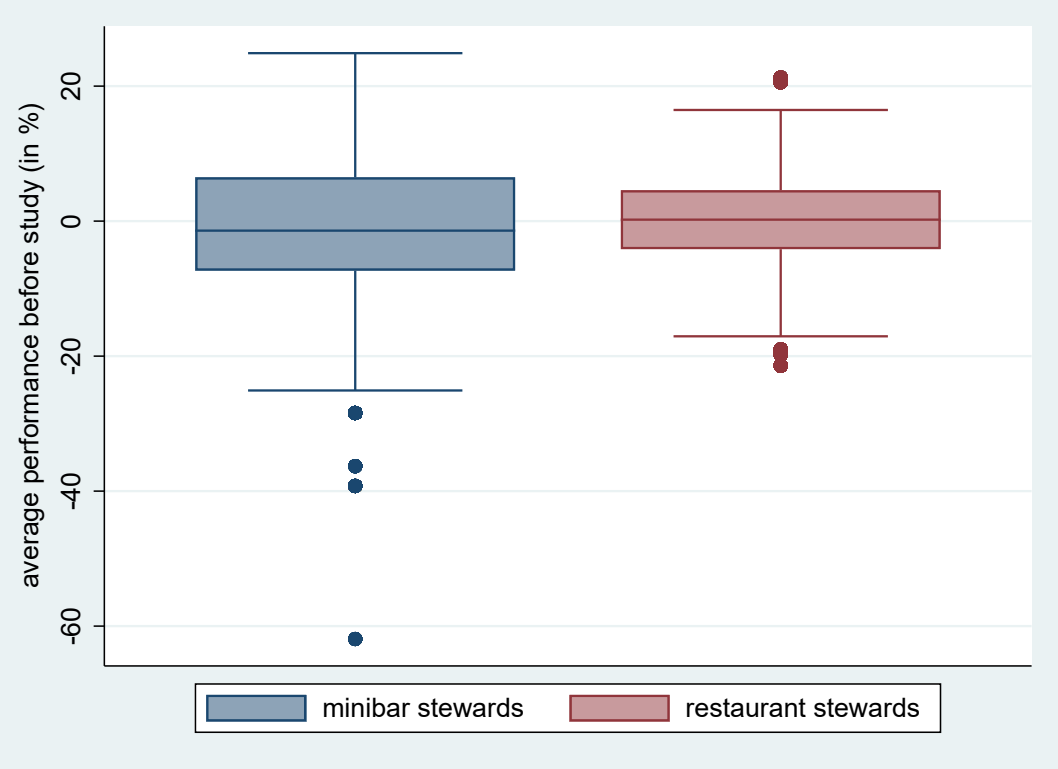

Figure 7: Performance differences between employees

Note: The average performance per steward shown above is calculated by the weighted mean of the monthly performance evaluations, as explained in Section 3.1 (weighted average over all months before the study period).

Appendix B Robustness check feedback effects (1)

Table 7: OLS regression log revenue per hour with clustered standard errors

	log revenue/hour
personal info	0.0174
	(0.0173)
social info	0.0329**
	(0.0129)
personal + social	0.0399**
	(0.0155)
performance before (in %)	0.0102***
	(0.0006)
service duration (in h)	-0.0141***
	(0.0032)
eating times (in %)	0.0076***
	(0.0007)
occupancy (in %)	0.0245***
	(0.0028)
occupancy^2	-0.0001***
	(0.0000)
1st class pass (in %)	0.0095***
	(0.0012)
No. stewards working	-0.7269***
	(0.1890)
event	0.1402***
	(0.0305)
weekend/holiday	0.1663***
	(0.0145)
Constant	3.4558***
	(0.2101)
shift type effects	Yes
city of shift start	Yes
month effects	Yes
Observations	6149
R2	0.3759

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix C Robustness checks feedback effects (2)

Table 8: Difference-in-Difference estimations (log revenue per hour)

	Diff-in-Diff ML regression	Diff-in-Diff FE regression	Diff-in-Diff FE without controls
personal info	-0.0004		
	(0.0127)		
social info	-0.0116		
	(0.0127)		
personal + social	0.0052		
	(0.0125)		
study	-0.0695***	-0.0697***	-0.0280***
	(0.0090)	(0.0113)	(0.0100)
personal info x study	0.0224*	0.0193	0.0090
	(0.0129)	(0.0178)	(0.0196)
social info x study	0.0439***	0.0411***	0.0453***
	(0.0117)	(0.0140)	(0.0150)
personal + social x study	0.0329***	0.0306*	0.0324*
	(0.0127)	(0.0173)	(0.0178)
performance before (in %)	0.0109***		
	(0.0005)		
service duration (in h)	-0.0037***	-0.0044*	
	(0.0011)	(0.0024)	
eating times (in %)	0.0055***	0.0054***	
	(0.0002)	(0.0005)	
occupancy (in %)	0.0167***	0.0167***	
	(0.0003)	(0.0005)	
occupancy^2	-0.0000***	-0.0000***	
	(0.0000)	(0.0000)	
1st class pass (in %)	0.0071***	0.0068***	
	(0.0005)	(0.0006)	
No. stewards working	-0.8896***	-0.8759***	
	(0.0401)	(0.1063)	
event	0.1449***	0.1477***	
	(0.0163)	(0.0232)	
weekend/holiday	0.1227***	0.1213***	
	(0.0054)	(0.0079)	
	(0.0012)		
Constant	3.7356***	3.6713***	3.9553***
	(0.0460)	(0.1153)	(0.0012)
shift type effects	Yes	Yes	No
city of shift start	Yes	Yes	No
month effects	Yes	Yes	No
sd beetween	0.0499	0.1534	0.1878
sd within	0.3083	0.3080	0.3563
Observations	32928	32928	33064

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix D Effect over time

Table 9: Treatment effects and time-interaction (log revenue per hour)

	interaction with study duration	interaction with dur + square	interaction with months
personal info x study duration	-0.0005	-0.0023*	
	(0.0004)	(0.0012)	
social info x study duration	0.0000	-0.0006	
	(0.0003)	(0.0011)	
personal + social x study duration	-0.0002	-0.0017	
	(0.0004)	(0.0014)	
personal info x duration2		0.0000	
		(0.0000)	
social info x duration2		0.0000	
		(0.0000)	
personal + social x duration2		0.0000	
		(0.0000)	
personal info x Apr			-0.0534*
			(0.0308)
personal info x May			-0.0397
			(0.0345)
personal info x Jun			-0.0497
			(0.0395)
social info x Apr			-0.0257
			(0.0290)
social info x May			0.0049
			(0.0313)
social info x Jun			-0.0075
			(0.0348)
personal + social x Apr			-0.0615*
			(0.0361)
personal + social x May			-0.0161
			(0.0388)
personal + social x Jun			-0.0316
			(0.0427)
fixed effects	yes	yes	yes
var(ID)	-3.1503***	-3.1667***	-3.1634***
	(0.1405)	(0.1400)	(0.1397)
var(Residual)	-1.2316***	-1.2332***	-1.2333***
	(0.0168)	(0.0167)	(0.0166)
Observations	6147	6147	6147

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix E DID margins

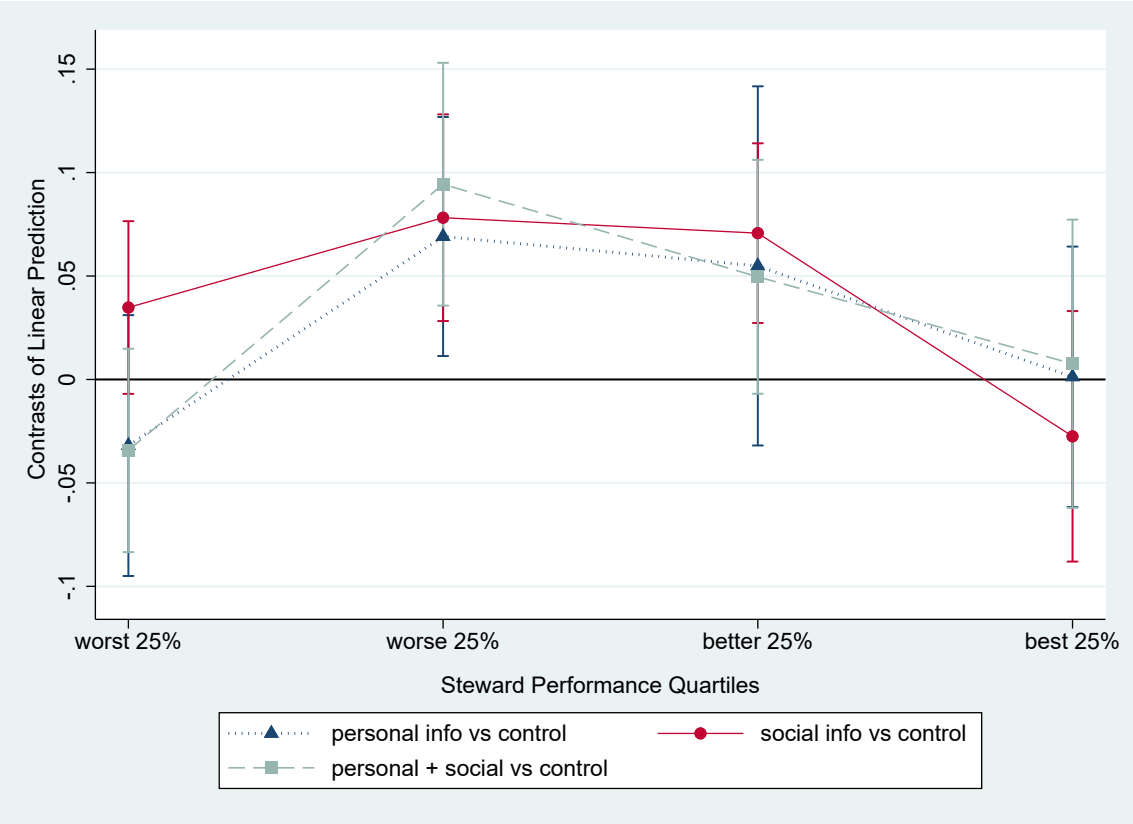

Figure 8: Contrasts of predictive margins DID with 95% Cis

Appendix F Immediate performance impacts 30 minutes

Table 10: Poisson Regression: Articles sold per 30 Minutes

	All treatment groups			Single treatments		
	(1)	(2)	(3)	Personal Info	Social Info	Persona + Social
feedback message	-0.021	-0.013	-0.006	0.032	-0.165**	-0.046
	(0.055)	(0.048)	(0.050)	(0.105)	(0.081)	(0.093)
working hour=4	0.106*	0.351***	0.499***	0.165	0.641***	0.490
	(0.056)	(0.049)	(0.132)	(0.204)	(0.237)	(0.324)
break (in min)	-0.078***	-0.086***	-0.101***	-0.097***	-0.102***	-0.127***
	(0.005)	(0.005)	(0.006)	(0.013)	(0.007)	(0.014)
occupancy (in %)	-0.002	0.010***	0.010***	0.013***	0.008***	0.008***
	(0.001)	(0.001)	(0.001)	(0.003)	(0.002)	(0.003)
Steward FE	No	Yes	Yes	Yes	Yes	Yes
Shift FE	No	Yes	Yes	Yes	Yes	Yes
Daytime FE	No	No	Yes	Yes	Yes	Yes
Month FE	No	No	Yes	Yes	Yes	Yes
Day of week FE	No	No	Yes	Yes	Yes	Yes
Pseudo R-sq	0.122	0.362	0.395	0.440	0.415	0.432
Observations	1986	1986	1986	578	753	655

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix G Immediate performance impacts in separate time frames

Table 11: Poisson Regression: Articles sold per working hour

	Working hour 3	Working hour 4
feedback message	-0.030	-0.092
	(0.040)	(0.056)
break (in min)	-0.030***	-0.026***
	(0.005)	(0.007)
occupancy (in %)	0.008***	0.001
	(0.001)	(0.001)
Steward FE	Yes	Yes
Shift FE	Yes	Yes
Daytime FE	Yes	Yes
Month FE	Yes	Yes
Day of week FE	Yes	Yes
Pseudo R-sq	0.510	0.529
Observations	997	597

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix H Immediate feedback-performance interactions

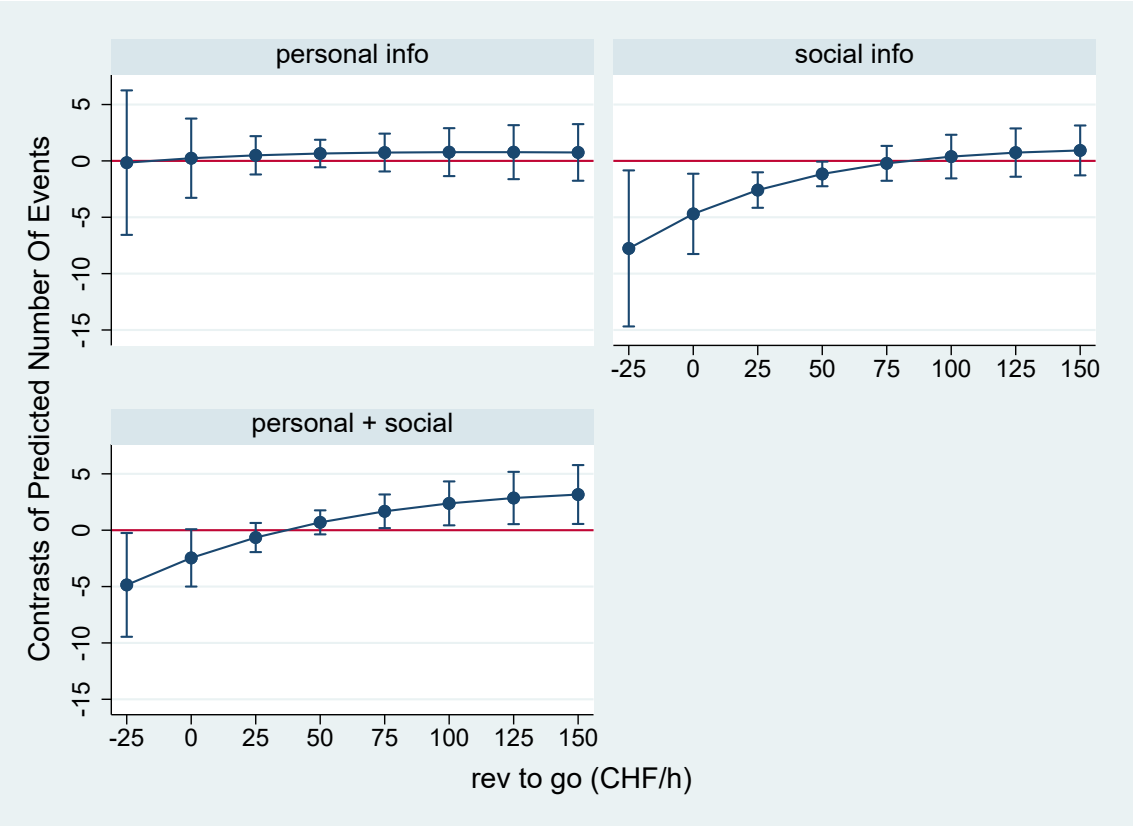

Figure 9: Contrasts of predictive margins with 95% Cis

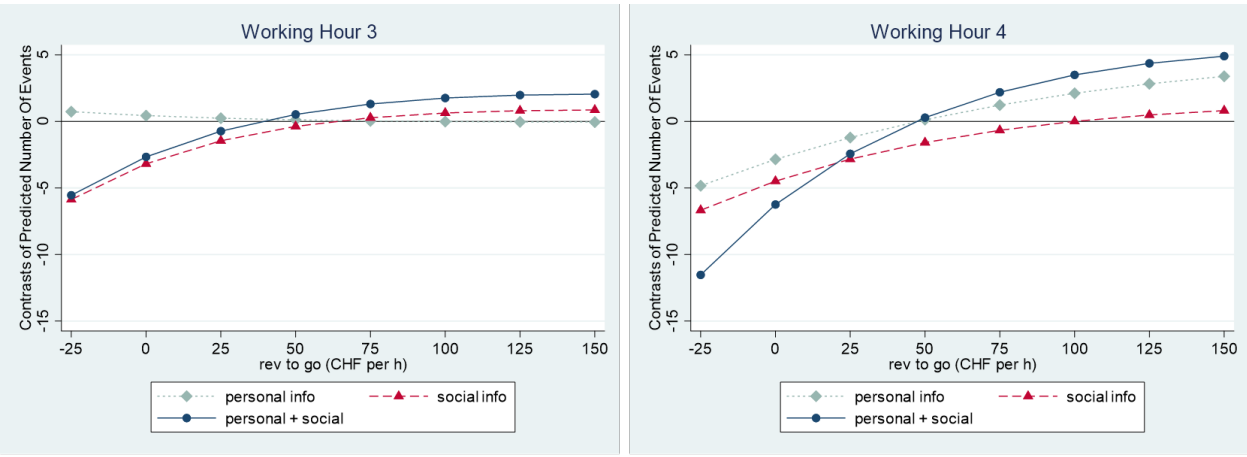

Figure 10: Contrasts of predictive margins for working hour 3 and 4

Appendix I Robustness checks immediate performance impacts

Table 12: Articles sold per working hour

	Negative Binomial Regression			Generalized Linear Mixed Model		
	(1)	(2)	(3)	(1)	(2)	(3)
feedback message	-0.251***	-0.217**	-0.223***	-0.653***	-0.453***	-0.375***
	(0.093)	(0.085)	(0.082)	(0.121)	(0.105)	(0.095)
rev to go	-0.003*	-0.010***	-0.009***	-0.011***	-0.012***	-0.010***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
feedback message x rev to go (CHF per h)	0.004	0.005**	0.004**	0.010***	0.008***	0.006***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
working hour=4	0.235***	0.097**	0.152**	0.219***	0.180***	0.222***
	(0.043)	(0.039)	(0.076)	(0.044)	(0.040)	(0.069)
break (in min)	-0.037***	-0.041***	-0.041***	-0.035***	-0.035***	-0.035***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
occupancy (in %)	0.001	0.006***	0.007***	0.006***	0.007***	0.009***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)
Service RE						
Steward FE	No	Yes	Yes	No	Yes	Yes
Shift FE	No	Yes	Yes	No	Yes	Yes
Daytime FE	No	No	Yes	No	No	Yes
Month FE	No	No	Yes	No	No	Yes
Day of week FE	No	No	Yes	No	No	Yes
Pseudo R-sq	0.031	0.099	0.114			
Observations	1994	1994	1994	1994	1994	1994

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$