Dynamic incentive effects of relative performance pay: A field experiment

Josse Delfgaauw a,b,⁎, Robert Dur a,b,c,d, Arjan Non c,e, Willem Verbeke a,f

a Erasmus University Rotterdam, The Netherlands
b Tinbergen Institute, The Netherlands
c CESifo, Germany
d IZA, Germany
e Maastricht University, The Netherlands
f ERIM, The Netherlands

HIGHLIGHTS

• We study dynamic incentive effects of relative performance pay.
• We conduct a large-scale field experiment among 189 stores of a retail chain.
• Our design creates exogenous variation in treatment stores’ relative performance.
• Treatment stores that lag far behind do not respond to relative incentives.
• Treatment stores close to winning a bonus do increase performance.

ABSTRACT

We conduct a field experiment among 189 stores of a retail chain to study dynamic incentive effects of relative performance pay. Employees in the randomly selected treatment stores could win a bonus by outperforming three comparable stores from the control group over the course of four weeks. Treatment stores received weekly feedback on relative performance. Control stores were kept unaware of their involvement, so that their performance generates exogenous variation in the relative performance of the treatment stores. As predicted by theory, we find that treatment stores that lag far behind do not respond to the incentives, while the responsiveness of treatment stores close to winning a bonus increases in relative performance. On average, the introduction of the relative performance pay scheme does not lead to higher performance.

1. Introduction

Non-linear pay-for-performance plans have dynamic incentive effects when employees receive intermediate performance information over the course of the incentive period. For instance, consider a salesman who can earn a bonus by attaining a monthly sales target while receiving daily or weekly sales figures. When realized sales during the month are such that it remains challenging but possible to reach the target, the bonus scheme provides strong incentives. The incentive effect is much weaker, however, when realized sales during the month are particularly high or low. High intermediate sales imply that the salesman can hardly miss the bonus, while low intermediate sales imply that the target is practically out of reach.

More generally, workers can use intermediate performance information to determine how much additional performance is necessary to obtain a bonus. This creates dynamic incentive effects, where the incentive effect of the pay-for-performance plan at each point in time depends on realized performance until then. Incentive plans based on relative performance, where prizes are awarded for outperforming sufficiently many competitors, are particularly prone to dynamic incentive effects. Sports leagues are a common example. In the workplace,
examples range from employee-of-the-month contests, to beat-the-index bonuses for stock brokers, and to job promotion contests.\(^1\)

Casas-Arce and Martinez-Jerez (2009) show formally that for contests with a large number of participants, the incentive effect of a relative performance incentive scheme is hump-shaped in lagged relative performance.\(^2\) Competitors who find themselves trailing far behind may perceive catching up to be impossible and consequently give up trying. Similarly, competitors who are far ahead may perceive losing as impossible and slack off as well. In contrast, incentives are highly salient for competitors who find themselves almost tied in intermediate performance and are at the margin of winning a prize. Analyzing sales contests among retailers of a commodities company, Casas-Arce and Martinez-Jerez (2009) find indeed that competitors in winning positions reduce performance when their lead increases. However, the performance of trailing competitors does not decrease when they lag further behind. The authors conjecture that this result might be affected by attrition bias. Frank and Obloj (2011) do find the predicted hump-shaped pattern in their analysis of a competition among units of a retail bank. Ludwig and Lünsen (2012) run a lab experiment to study the role of intermediate relative performance information in tournaments, and also find that leading competitors slack off, while trailing competitors increase their stated efforts.\(^3\)

Testing for the presence and strength of dynamic incentive effects is hampered by two issues. First, in contests with a limited number of participants, a competitor’s optimal strategy depends on (its perception of) its competitors’ strategies. For example, a trailing competitor may be best off by accepting its loss when the other competitors keep effort high, but not when they would slack off. Second, serial correlation in performance biases the estimates of the effect of intermediate relative performance on subsequent performance. For instance, positive serial correlation would imply that a positive shock to performance in the previous period increases both relative intermediate performance and current performance. Casas-Arce and Martinez-Jerez (2009) and Frank and Obloj (2011) employ a method developed by Arellano and Bond (1991) that relies on taking first differences and using lagged values of independent variables as instruments for the independent variables to correct for this bias.

In this paper, we take a unique approach in tackling both issues by setting up a relative performance pay scheme where only one of the ‘competitors’ can earn a prize, while the other participants are kept unaware of their involvement. This implies that the strategies of all non-competing participants are exogenous, allowing us to use their performance as an instrument for intermediate relative performance of the competing participant. More specifically, we study the dynamic incentive effects of this relative performance pay scheme by conducting a natural field experiment in a Dutch retail chain. We provide the employees of 93 stores randomly selected from 189 of the company’s stores with the opportunity to earn a bonus. The bonus is awarded when a treatment store outperforms three comparable stores from the control condition over the course of a four-week period (February 2010). Each week, treatment stores receive a poster with the intermediate performance of all four stores in their group. Importantly, the employees of the three comparison stores do not receive the poster, cannot earn a bonus, and do not learn that another store can earn a bonus by beating their performance. This way the treatment stores compete against stores that are not competing, and treatment stores were informed about this.

This setup has two advantages. First, we can use the performance of the three comparison stores as an instrument for trailing behind or being ahead: lagged performance of the comparison stores does affect intermediate relative performance, but does not affect current performance of the treatment store other than through lagged relative performance. Hence, using this instrument, our estimates are not biased by serial correlation in stores’ own performance. Moreover, we eliminate the possible influence of perceptions of competitors’ strategies from the estimations. This makes for a clean test of dynamic incentive effects. A disadvantage of our setup is that the estimated effects of our competition may not generalize to settings where all contestants compete. Thus, while our experiment cleanly tests whether the competing stores respond to intermediate relative performance feedback as predicted by theory, it may not yield a reliable estimate of the overall effect of standard tournaments on performance. In the terminology of Ludwig et al. (2011), we conduct a mechanism experiment.

Our results are as follows. First, we find a positive effect of intermediate relative performance on current performance for stores close to the target, particularly in the last two weeks of the experiment. This effect is substantial: a one percentage point increase in intermediate relative performance increases current performance by 0.73%. Stores lagging far behind do not respond to intermediate relative performance. This suggests that the employees in these stores gave up trying to win. Hence, as predicted by theory, we find that changes in intermediate relative performance matter more for competitors that perform close to the target than for competitors that lag far behind. During the contest, hardly any treatment store managed to get far ahead of all its comparison stores. Hence, we cannot test the hypothesis that high-performers slack off as their lead increases.

Second, we find no average treatment effect of introducing the contest, neither for the four weeks taken together nor for one of the weeks separately. This contrasts with several recent findings on the incentive effects of tournaments. In another retail chain, we do find a substantial positive effect of introducing a standard tournament among shops (Delfgaauw et al., 2013), as do Erev et al. (1993) and Bandiera et al. (2013) among teams of fruit pickers and Casas-Arce and Martinez-Jerez (2009) among retailers of a commodities company. Even more striking, several recent papers suggest that the mere provision of relative performance feedback can be sufficient to trigger higher performance (Azmat and Iriberri, 2010; Blanes i Vidal and Nossol, 2011; Delfgaauw et al., 2013, and Kosfeld and Necker-\text{m}ann, 2011), Bandiera et al. (2013) and Barankay (2012) obtain an opposite result. A possible explanation for the lack of a significant average treatment effect in our experiment is that only a limited number of stores happened to be close to the target for winning, while the majority of stores lagged far behind. As a result, the number of stores positively responding to the contest is just too small to be reflected in a significant average treatment effect. An alternative explanation is that beating unaware contestants, as in our setting, is less exciting than beating competing contestants. Relatedly, participants in our benchmark competitions may anticipate weaker feelings of envy after losing as compared to participants in regular tournaments (Eisenkopf and Teys\text{-}sier, 2013).

Our experiment also relates to a literature on the incentive effects of non-linear payment schemes. Forbes et al. (2012) find that airline personnel who are rewarded for on-time performance reduce taxi-in times only when the expected arrival time is just around the critical threshold for the flight being recorded as ‘late’. Schweitzer et al. (2004) and Cadsby et al. (2010) find in lab experiments that non-linear incentive schemes invite substantial lying to meet the target. These findings are well in line with ours in the sense that individuals respond to performance feedback if they are sufficiently close to the target. An important difference with our study is that we aim to identify

---

1. As a concrete example, more than half of the remuneration of the executive directors of oil company Shell is based on a ranking of Shell’s performance relative to its four main competitors on four publicly available measures. The incentive plan has a three-year horizon, during which the companies regularly release the latest figures with respect to these performance measures (Royal Dutch Shell, 2009).

2. Following the early literature on tournament theory (Lazear and Rosen, 1981; Green and Stokey, 1983; Nalebuff and Stiglitz, 1983), most of the literature has abstracted from dynamic incentive effects of tournaments. Recently, a string of theoretical papers has studied the cost and benefit to a principal of providing intermediate relative performance feedback to his agents (Gershovik and Perry, 2009; Asaygi, 2010; Ederer, 2010; Goltsman and Mukherjee, 2011).

3. Relatedly, Fershtman and Gneezy (2011) let kids run side-by-side and find that increasing incentives yield higher performance but also a higher fraction of kids giving up during the race.
Intermediate ranking after 2 weeks:

1. R.dam/Lijndaan  + 9 %
2. Schiedam  + 7 %
3. Spijkenisse  + 6 %
4. Utrecht/H.C.  + 2 %

Percentage is the difference between sales and budgeted sales over February up to now.

Fig. 1. Example of the weekly feedback posters (translated into English).

increases in sales performance rather than artificial improvements in recorded performance or outright lying.

Our experiment involves one incentive period of four weeks. When incentive schemes are repeated over time, as with monthly or year-on-year targets, other types of dynamic incentive effects may arise. For instance, sales may be shifted forward or backward in time around the incentive commencement date in order to meet the current target or to alleviate the difficulty of meeting the next target; see Asch (1990), Oyer (1998), Courty and Marschke (2004), and Larkin (2014) for empirical evidence. Furthermore, when the targets in repeated incentive schemes are based on historical performance, workers have an incentive to beat the target by only a limited amount even it would be possible to greatly outperform the target. Bouwens and Kroos (2011) find evidence in line with such ratchet effects, using store-level data from a retail chain. Cooper et al. (1999) and Charness et al. (2010) find ratchet effects in the lab. Ratchet effect considerations may be yet another explanation for why we find no average treatment effect, as workers may have feared that a strong response to the introduction of the relative performance pay scheme would result in higher targets in their regular incentive scheme.

We proceed as follows. In the next section, we describe the context and design of our experiment in detail. Then, in Section 3, we describe the econometric model and estimation strategy. Section 4 presents the results of the estimations and a number of robustness checks. Finally, Section 5 concludes.

2. Experimental context and design

2.1. Experimental set-up

The experiment took place in February 2010 in a retail chain in The Netherlands that sells computer games, music, and movies. At the start of 2010, the retail chain owned 208 geographically dispersed stores, operating under two different brands. Each store employs on average 5 employees, including a store manager. The company’s central management decides on the range of products sold, pricing, and advertisement. New products arrive in stores complete with instructions on how to sell them. Store managers are responsible for day-to-day operations.

In this environment, store employees may have limited scope to affect sales. Still, the company’s management is convinced that employees can contribute to sales, in particular through cross-selling. The company instructs employees to show interest in potential customers, to help and give advice whenever possible, and to suggest related products. Employees receive rather weak incentive pay on top of their base salary, based on their shop’s yearly sales growth and a subjective performance evaluation. The company’s management was not fully satisfied with this incentive scheme and wished to learn more about the effects of short-term incentives, in particular of sales contests. The pre-existing incentive scheme remained in place during the experiment. Hence, even though incentives may not have very large effects in this retail chain, the company’s management sees sufficient scope for incentives to have a beneficial effect on performance.

We designed a relative performance incentive scheme to be implemented in a randomly selected subset of stores (the treatment condition), while the rest of the stores comprised the control condition. All employees (including the shop manager) of a store in the treatment condition could earn a bonus by outperforming three preselected stores from the control condition by a sufficient margin. Stores in the control condition could not earn a bonus, and employees in the treatment stores were informed about this. Performance is measured as cumulative sales revenue in percentage deviation of budgeted sales in February 2010 (a period of 4 weeks). Budgeted sales are set in advance by the company’s central management and cannot be affected by stores. Let \( Y_{sw} \) be sales and \( b_{sw} \) budgeted sales of store \( s \) in week \( w \), respectively. Weekly performance \( p_{sw} \) is given by

\[
p_{sw} = \frac{y_{sw} - b_{sw}}{b_{sw}} \cdot 100\%
\]  

(1)

and cumulative performance over February 2010 is given by

\[
p_{C} = \frac{\sum_{w=E1}^{E4} (y_{sw} - b_{sw})}{\sum_{w=E1}^{E4} b_{sw}} \cdot 100\%
\]  

(2)

where the summation is over the four experimental weeks \( E1–E4 \), namely week 5, 2010 to week 8, 2010. Below, we will refer to cumulative performance during the experimental weeks as performance in the tournament.

All employees of a treatment store received a bonus of gross 150 euro each when their shop’s performance in February 2010 was at least 10 percentage points higher than the performance of all three comparison stores. When a treatment store scored between 5 and 10 percentage points above the performance of all three comparison stores, the store manager (including the shop manager) of a store in the treatment condition could earn a bonus by outperforming three preselected stores from the control condition by a sufficient margin. Stores in the control condition could not earn a bonus, and employees in the treatment stores were informed about this. Performance is measured as cumulative sales revenue in percentage deviation of budgeted sales in February 2010 (a period of 4 weeks). Budgeted sales are set in advance by the company’s central management and cannot be affected by stores. Let \( Y_{sw} \) be sales and \( b_{sw} \) budgeted sales of store \( s \) in week \( w \), respectively. Weekly performance \( p_{sw} \) is given by

\[
p_{sw} = \frac{y_{sw} - b_{sw}}{b_{sw}} \cdot 100\%
\]  

(1)

and cumulative performance over February 2010 is given by

\[
p_{C} = \frac{\sum_{w=E1}^{E4} (y_{sw} - b_{sw})}{\sum_{w=E1}^{E4} b_{sw}} \cdot 100\%
\]  

(2)

where the summation is over the four experimental weeks \( E1–E4 \), namely week 5, 2010 to week 8, 2010. Below, we will refer to cumulative performance during the experimental weeks as performance in the tournament.

All employees of a treatment store received a bonus of gross 150 euro each when their shop’s performance in February 2010 was at least 10 percentage points higher than the performance of all three comparison stores. When a treatment store scored between 5 and 10 percentage points above the performance of all three comparison stores, all of its employees
received 75 euro. Lastly, outperforming all three comparison stores by less than 5 percentage points yielded a cake for the treatment store, but only if the treatment store also performed above budget.6

All communication on the experiment towards the shops went through the company's regular channels, so shop managers and employees were not aware of our involvement. Hence, our experiment classifies as a natural field experiment (Harrison and List, 2004). In January 2010, the company informed all store managers and employees that a randomly selected set of stores would get the opportunity to earn a bonus in February 2010, and that all other stores would have a similar opportunity later that year. On January 22, all store managers and employees in the treatment stores were informed about the details of the relative performance incentive scheme, which would start on Monday February 1. At this point, we did not inform treatment stores with whom they were matched. Control stores were not informed about their role in the treatment stores’ incentive scheme.

During the experiment, we provided weekly feedback to the treatment stores on their relative performance in the form of a poster. The poster contained the cumulative sales relative to budget figures of the treatment shop and its three comparison shops, ranked in descending order. In order to ensure credibility, we published the identity of the comparison stores along with their performance figures, see Fig. 1 for an example. Thus, the identity of the comparison stores was revealed when the first poster arrived, on February 9. Furthermore, in the first week of the experiment, all treatment stores received a large poster, with room to glue on the four posters with weekly rankings to be received in the following weeks. We created the feedback posters and sent them in the company's envelopes by regular mail to the stores. Store managers were instructed to put up these posters in the store's canteen.9 Stores in the control condition did not receive posters.

It is possible that control stores learned about the details of the experiment, or that treatment stores contacted their comparison stores after receiving the first feedback poster. According to the central management staff, normally there is some communication between stores, in particular between store managers within a region. To reduce possibilities of collusion, treatment stores were never matched with another store from their region (the assignment procedure is discussed in detail in the next subsection). For control stores, engaging in collusive actions is not attractive, as it reduces their regular incentive pay. During the experiment, central management staff did not receive questions about the incentive event from control stores, nor did they hear of any treatment store contacting their comparison stores. Hence, we are quite confident that control stores were not aware of the details of the experiment.10

Our design has two advantages as compared to a regular competition. First, as treatment stores only receive a bonus when they outperform comparable stores from the control condition, the payout is relatively low when the incentive has little effect on performance. This was seen as a major benefit by the company's management. Second, performance of the comparison stores is exogenous to the incentive scheme, as these stores could neither earn a bonus, nor received relative performance feedback, and were not aware that their performance played a role in the incentive scheme. We exploit differences in comparison stores’ performance during the experiment to analyze how treatment stores’ intermediate relative performance affects the effect of the incentive scheme in subsequent weeks.

2.2. Assignment procedure

The aim of our assignment procedure is to match stores from the treatment condition to similar stores from the control condition. It is important to create homogeneous groups of stores, as theory predicts that differences in ability between contestants weaken incentive effects of tournaments (Lazear and Rosen, 1981; O’Keefe et al., 1984; Rosen, 1986; see Fonseca, 2009 and Höchtl et al., 2011 for empirical evidence). We therefore use weekly sales and budget data for the weeks 40 to 53 in 2009 to match stores on the basis of their historical performance. As the company's management excluded a specific group of 14 stores from participating in the experiment, 194 of the company’s 208 stores were included in the matching procedure.10 Besides historical performance, the company's management argued that store size was an important characteristic to take into account when matching treatment stores with comparison stores, as employees in small treatment stores might perceive it as unfair when matched to large comparison stores or vice versa (e.g. because of differences in local demand conditions, quality of management, free rider effects, and so on). Therefore, we first created four equally large strata based on store size as measured by average weekly sales revenues. Randomly, half of the stores in each stratum was assigned to the treatment condition, while the remaining half of the stores were assigned to the control condition. Subsequently, we matched each treatment store to three control stores from the same stratum. Our randomization procedure ruled out that, by chance, there would be relatively many treatment stores in a particular stratum, which could impede the creation of a level playing field for treatment stores in this stratum (as there would be few control stores with similar past performance and similar store size left). To reduce opportunities for collusion, we imposed that each treatment store was matched to control stores located in other regions, as discussed in the previous

---

5 For employees who did not have a full-time contract, the size of the bonus was proportional to the contractual number of hours. Hourly wages are close to the minimum wage, which makes that receiving the high bonus would increase monthly earnings by about 10%.

6 The latter requirement only applied for the cake, not for any of the two bonuses. This requirement was a last-minute addition by the company's management to the rules.

7 We organized two-stage elimination tournaments in the Fall of 2010, see Delfgaauw et al. (2014).

8 The company's regional managers were instructed to verify that store managers actually put up the posters in the canteen. We have not heard about a single store manager who refused to do so.

9 Another potential source of contamination of the control group is relocation of employees or store managers from treatment stores to control stores. We have no information on the frequency of such relocations. However, relocations typically take place at the beginning or end of the month, so when stores were not yet informed on the identity of their competitors. Therefore, and because our experiment lasts only four weeks, we don't think this could possibly influence our results.

10 The group that is excluded from participation consists of all stores that are located in railway stations. The company's management considered those stores as special cases. As will be explained below, we had to drop another 5 stores from the experiment, leaving us with a final sample of 185 stores.
and friends, much like Christmas in many other countries. The essence of the festivities is an evening of gift-giving among relatives. The company distinguishes between 12 geographically-clustered regions, each led by a different regional manager. We used the same region classification in our matching procedure. Apart from this regional separation, treatment stores were matched to the control stores that were most comparable in terms of the performance measure (cumulative sales revenue relative to the budget) for the period of week 40 to week 53 in 2009. Note that a control store can be matched to multiple treatment stores. After this assignment procedure, we excluded one treatment store from the experiment as its budget figures were unavailable during the experimental period, which made it impossible to determine performance in the tournament. Furthermore, 3 treatment stores and 1 control store were shut down in January 2010. This leaves us with 189 stores in the experiment: 93 stores in the treatment condition and 96 stores in the control condition. Lastly, in week 7 of 2010, a total of 29 stores were closed for one or two days in relation to carnival festivities, mainly in the south of The Netherlands. Treatment stores were slightly less often closed than control stores, but not significantly so. In all estimations below, we correct for the effect of carnival.

As a first hint of the overall effect of the relative performance incentive, Table 1 shows that there is no difference in average sales and performance between treatment and control stores for the treatment weeks (week 5, 2010 to week 8, 2010). Figs. 4 and 5 provide further insight into the overall treatment effect, by plotting the differences between treatment and control stores regarding the within-store standard deviation of sales and performance. Further, the number of employees per store does not differ significantly between the treatment and control stores. Lastly, in week 7 of 2010, a total of 29 stores were closed for one or two days in relation to carnival festivities, mainly in the south of The Netherlands. Treatment stores were slightly less often closed than control stores, but not significantly so. In all estimations below, we correct for the effect of carnival.

Table 1

<table>
<thead>
<tr>
<th></th>
<th>All stores</th>
<th>Treatment stores</th>
<th>Control stores</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sales</strong></td>
<td>Mean</td>
<td>Std</td>
<td>Mean</td>
</tr>
<tr>
<td>Sales weeks 40/2009–53/2009</td>
<td>100.00</td>
<td>40.31</td>
<td>100.04</td>
</tr>
<tr>
<td>Sales weeks 5/2010–8/2010</td>
<td>115.18</td>
<td>46.62</td>
<td>115.27</td>
</tr>
<tr>
<td>Budgeted sales</td>
<td>71.67</td>
<td>29.90</td>
<td>71.56</td>
</tr>
<tr>
<td>Budgeted sales weeks 40/2009–53/2009</td>
<td>104.56</td>
<td>41.54</td>
<td>105.33</td>
</tr>
<tr>
<td>Performance</td>
<td>−0.03</td>
<td>0.12</td>
<td>−0.04</td>
</tr>
<tr>
<td>Performance weeks 40/2009–53/2009</td>
<td>−0.06</td>
<td>0.11</td>
<td>−0.07</td>
</tr>
<tr>
<td>Average within-store standard deviation sales</td>
<td>44.27</td>
<td>21.06</td>
<td>44.40</td>
</tr>
<tr>
<td>Av. within-store st. dev. sales weeks 40–53/2009</td>
<td>49.18</td>
<td>24.16</td>
<td>49.33</td>
</tr>
<tr>
<td>Average within-store st. dev. performance</td>
<td>0.19</td>
<td>0.12</td>
<td>0.18</td>
</tr>
<tr>
<td>Av. within-store st. dev. performance weeks 40–53/2009</td>
<td>0.18</td>
<td>0.16</td>
<td>0.17</td>
</tr>
<tr>
<td>Av. within-store st. dev. performance weeks 5–8/2009</td>
<td>0.11</td>
<td>0.05</td>
<td>0.10</td>
</tr>
<tr>
<td>Average within-store st. dev. performance corrected for common shocks</td>
<td>0.14</td>
<td>0.13</td>
<td>0.13</td>
</tr>
<tr>
<td>Number of stores</td>
<td>189</td>
<td>93</td>
<td>96</td>
</tr>
</tbody>
</table>

Performance is defined as (sales-budgeted sales)/budgeted sales. For confidentiality reasons, sales and budgeted sales figures are indexed to the average sales per store per week over the whole sample. None of the differences between treatment stores and control stores are significant at the 10%-level.

11 The decision to terminate these stores has been made before we conducted the randomization, but was not communicated to us. The closure of these stores is therefore not related to assignment to the treatment condition. Moreover, these stores were already closed before January 22, when we informed the stores about the experiment. Hence, possible relocations of personnel from closed treatment stores to control stores took place before we communicated the details of the experiment to the treatment stores.

12 Sinterklaas takes place the fifth of December and is widely celebrated in the Netherlands. The essence of the festivities is an evening of gift-giving among relatives and friends, much like Christmas in many other countries.

13 To be exact, 23, 22, 21, and 19 stores were in the leading position at the end of tournament weeks 1, 2, 3, and 4, respectively. Absent treatment effects, the expected number is 93/4 = 23.25.
average, treatment stores lagged 11% behind their best comparison store. In about 25% of all cases, stores were informed that they lagged more than 20% behind, while 13.6% of the stores were at some point more than 5% ahead of their best competitor. Stores were rarely more than 10% ahead: this happened in only 6% of all cases.

3. Method

In this section, we describe our empirical strategy. We estimate the effects on sales rather than on the performance measure used in the incentive scheme as described by Eq. (1). The reason is that budgeted sales are set in advance by the company’s central management. Hence, shops can affect their performance only through sales. Estimation results for sales and performance as measured by Eq. (1) are therefore similar. Since effects on sales are more easy to interpret, we use sales rather than performance as our main outcome measure. We include 22 weeks in the estimations, where the final four weeks are the experimental weeks (February 2010). We assess the average effect of the relative performance incentive scheme on sales using OLS with week-fixed effects and store-fixed effects. That is, we estimate

\[
\ln(y_{s,w}) = \alpha_s + \theta_w + \gamma B_{s,w} + \kappa F_{s,w} + \epsilon_{s,w}
\]

where \(\ln(y_{s,w})\) is the log of sales of store \(s\) in week \(w\).\(^{14}\) Store and week-fixed effects are given by \(\alpha_s\) and \(\theta_w\), respectively. \(B_{s,w}\) is a dummy variable that takes the value one from week 5 to week 8 in 2010 for stores in the treatment condition, and is zero otherwise. \(F_{s,w}\) measures the number of days shop \(s\) is closed for carnival festivities in week \(w\) (this variable takes positive values only in week 7, 2010), and \(\epsilon_{s,w}\) is an error term, possibly serially correlated (i.e. \(\tilde{E}(\epsilon_{s,w};\epsilon_{s,w} - 1) \neq 0\)).

The main goal of this paper is to analyze the effect of intermediate relative performance on subsequent performance. First, we introduce some additional notation. Let \(T\) and \(C\) be the sets of stores in the treatment and the control condition, respectively. Further, denote by \(C_t \subseteq C\) a control store matched to treatment store \(t \in T\). Lastly, let \(p^{CU}_{t,w} = 1\) denote the cumulative performance of store \(s\) during the experiment up to but not including week \(w\), as measured by cumulative sales over budget in February 2010 (i.e. weeks 5 to 8 in 2010, which corresponds to weeks 19–22 in our dataset):

\[
p^{CU}_{t,w-1} = \begin{cases} 
\frac{\sum_{w=10}^{w-1} (y_{s,w} - B_{s,w})}{\sum_{w=10}^{w-1} B_{s,w}} & \text{100% if } w \in [20, 21, 22] \\
0 & \text{if } w \not\in [20, 21, 22]
\end{cases}
\]

Hence, \(p^{CU}_{t,w-1}\) is the performance figure for store \(s\) as depicted on the poster received at the start of week \(w\) during the experiment. The effect of intermediate performance of treatment stores relative to their best-performing comparison store on subsequent sales can be estimated by

\[
\ln(y_{s,w}) = \alpha_s + \theta_w + \gamma B_{s,w} + \kappa F_{s,w} + \epsilon_{s,w}
\]

where the term \(p^{CU}_{t,w-1} - \max_{t \in T} (p^{CU}_{t,w-1})\) gives the difference in cumulative performance during the experiment between treatment store \(t\) and its best-performing comparison store \(s\), up to and including the previous week.\(^{15}\) Since the experiment lasted four weeks, we have three intermediate relative performance figures per treatment store, corresponding to a total of 279 treatment store–week observations.

The estimation of the effect of intermediate relative performance on subsequent sales, represented by \(\mu\) in Eq. (5), is biased in case sales are serially correlated. When \(\epsilon_{s,w}\) is correlated with \(\epsilon_{s,w - 1}\), cumulative past performance \(p^{CU}_{t,w - 1}\) is no longer exogenous to current sales \(y_{s,w}\). This can be seen from Eq. (4); \(p^{CU}_{t,w - 1}\) is a function of \(y_{s,w - 1}\), which is correlated with \(y_{s,w}\) via the serial correlation in the error structure.\(^{16}\) Thus, serial correlation in treatment stores’ own performance leads to biased estimates of the parameter of interest \(\mu\). By construction, past performance of the best control store matched to treatment store \(t\), \(\max_{t \in T} (p^{CU}_{t,w - 1})\), is exogenous to the treatment store’s sales \(y_{s,w}\).\(^{17}\) Control stores do not earn a bonus and do not receive feedback, implying that their past performance only influences the treatment stores’ current performance via the feedback provided to the treatment stores. Therefore, we instrument the

\(^{14}\) The reason for using the log of sales rather than sales is that, theoretically, we would expect that treatment effects (as well as week fixed-effects) influence sales for each store with the same percentage, not with the same absolute amount (as a given amount may be little for large stores, but a lot for small stores).

\(^{15}\) This term is set to zero for control stores.

\(^{16}\) Note that \(p^{CU}_{t,w - 1}\) is also endogenous when \(\epsilon_{s,w}\) is correlated with \(\epsilon_{s,w - 2}\) or \(\epsilon_{s,w - 3}\), as \(p^{CU}_{t,w - 1}\) also includes these longer lags in weeks 21 and 22, respectively.

\(^{17}\) Exogeneity of control stores’ performance is violated when there are region-specific shocks that are correlated both across regions and over time. As a robustness check, we will also estimate Eq. (5) including region-specific week-fixed effects, see Section 4.2.
The expected cumulative performance of treatment store $i$ in the experiment, $E[p_{i,w-1}^{CU}]$, is based on the average performance prior to the start of the experiment, $\alpha_w$, and the realized performance of the common shock $\theta_{w-1}$. As week-fixed effects by definition show up in the best comparison store’s fixed effects, we cannot use the actual difference between the lagged cumulative performance and the expected difference, $D_{i,w-1}$, for the last three weeks of the experiment. Clearly, there exists a strong association between the two (the correlation coefficient is $0.65$).

Eq. (5) estimates a linear effect of intermediate relative performance. However, the incentive scheme is likely to have the biggest effect when treatment stores learn that they are close to the relative performance targets for winning a bonus (Casas-Arce and Martinez-Jerez, 2009). Treatment stores lagging far behind in the intermediate ranking may give up, and treatment stores far ahead may reduce their efforts when they anticipate that they can hardly miss the bonus. As we have seen in the previous section, in the course of the experiment, we have many treatment stores that face an uphill battle, while there are only few stores that are comfortably ahead. In total, we have only 8 store-week observations where treatment stores’ intermediate relative performance is more than 10 percentage points above the target for the high bonus (i.e. with $p_{i,w-1}^{CU} - \max_{c} [p_{c,w-1}^{CU}] > 20\%$). This implies that we cannot test whether stores that greatly outperform their comparison stores reduce their efforts. We can test whether the marginal effect of intermediate relative performance on current performance differs between stores that lag far behind and stores that are still in the running, by allowing the estimated effect to differ between both groups.

In determining which stores still have a chance of earning a bonus, we cannot use the actual difference between the lagged cumulative performance of the treatment store and its best control, as given by Eq. (4). Serial correlation in $p_{i,w}$ would bias the estimates. Hence, we again use the estimated difference Eq. (6) to determine stores’ chances of earning a bonus. Rather arbitrarily, we set the bar for being too far behind at a 5 percentage point lag relative to the best performing comparison store. Note that stores that lag 5 percentage points behind need to improve their relative performance by 5 percentage points in order to win a cake and by at least 10 percentage points to obtain a bonus. We do vary the bar to assess the robustness of the results. Let $l_{i,w-1}$ be a dummy that takes value 1 during experimental weeks for treatment stores whenever $D_{i,w-1} > 0$.

---

18 We weight the week-fixed effects by budgeted sales $b_{i,w}$ to account for the fact that weeks with a higher absolute budgeted sales volume have a higher weight in cumulative performance, see Eq. (4).

19 The theory developed by Casas-Arce and Martinez-Jerez (2009) predicts that performance is hump-shaped in intermediate relative performance, but does not predict the exact level of relative performance at which the incentive effect peaks.

20 Excluding these 8 observations from the analysis does not affect any of the results.
again instrumenting the difference in intermediate performance $p_{t,w}^{CU} - \max[p_{t,w-1}^{CU}]$ by the expected difference $D_{t,w} - 1$ as given by Eq. (6).\(^{21}\) Note that, as compared to Eq. (5), Eq. (8) allows for a differential treatment effect for stores close to winning as well as for a different effect of intermediate relative performance for stores close to winning.

Treatment stores’ performance may, in addition to the distance to the best control store, also depend on the distance to the second-best control store. Unfortunately, we cannot disentangle the two effects due to problems of multicollinearity. As it turns out, the correlation between distance to second best control store and best control store also depend on the distance to the second-best control store. Unfortunately, we cannot disentangle the two effects due to problems of multicollinearity. As it turns out, the correlation between distance to second best control store and best control store also depend on the distance to the second-best control store.

In all of our estimations, we cluster standard errors at the store level to correct for serial correlation within stores, as well as for heteroscedasticity across stores (see Bertrand et al. (2004) for a discussion of the importance of correcting for serial correlation in Difference-in-Difference estimation).

4. Results

4.1. Estimation results

This subsection presents the estimation results. We subsequently present our estimates of the average treatment effect, the average treatment effect for stores close to winning, and how these treatment effects depend on interim relative performance. In the next subsection, we present the results of a number of robustness checks we conducted.

The estimates of the average treatment effect are presented in Table 2. The first column in Table 2 gives the results of estimating Eq. (3). On average, the relative performance incentive scheme did not significantly affect sales. This result is not due to a lack of statistical power: given the size of the estimated standard errors, we should be able to detect reasonably small effect sizes. However, the point estimate is also very close to zero. The second column of Table 2 shows that there is some variation in the estimated treatment effect by week, but none of

\(\text{Table 2. Average treatment effect.} \)

<table>
<thead>
<tr>
<th>Dependent variable: $\ln(\text{sales})$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>−0.004</td>
<td>−0.008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment + close</td>
<td>0.037</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment week 1</td>
<td>−0.003</td>
<td>−0.003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment week 2</td>
<td>−0.013</td>
<td>−0.025</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment week 3</td>
<td>−0.020</td>
<td>−0.022</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment week 4</td>
<td>0.021</td>
<td>0.015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment week 2 + close</td>
<td>0.056*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment week 3 + close</td>
<td>0.017</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment week 4 + close</td>
<td>0.039</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carnival</td>
<td>−0.026*</td>
<td>−0.028*</td>
<td>−0.026*</td>
<td>−0.028*</td>
</tr>
<tr>
<td>Store-fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Week-fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Store–week observations</td>
<td>4158</td>
<td>4158</td>
<td>4158</td>
<td>4158</td>
</tr>
<tr>
<td>Stores</td>
<td>189</td>
<td>189</td>
<td>189</td>
<td>189</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.9281</td>
<td>0.9281</td>
<td>0.9281</td>
<td>0.9281</td>
</tr>
</tbody>
</table>

Standard errors clustered at the store level in parentheses.

"Close" is a dummy variable that takes value one when the store’s expected intermediate performance is at most 5 percentage points below its best comparison store, i.e. when $D_{t,w} - 1 > -5\%$ (see Eq. (6)).

Standard errors clustered at the store level in parentheses.

\(^{21}\) Instead of estimating Eq. (8), we could estimate a quadratic specification of intermediate relative performance. However, the estimates for the quadratic specification would be heavily affected by the many treatment store–week observations with sizable negative intermediate relative performance (see Fig. 6). Hence, we would learn little about the marginal effect of intermediate relative performance for stores close to winning a bonus.
the estimates differs significantly from zero. This suggests that the absence of a treatment effect is not due to stores gradually becoming discouraged, which would imply a negative trend in the estimated treatment effects, or learning to improve, which would imply an upward trend.

The main aim of our analysis is to estimate how the treatment effect depends on intermediate relative performance. As OLS estimation is possibly biased by serial correlation in the error term, we use the predicted difference, $\bar{D}_{t, w - 1}$ as defined by Eq. (6), as an instrument for the actual difference in cumulative performance between the treatment stores and their best comparison stores, $p_{CU t, w - 1} - \max_{i} p_{CU t, w - 1}$. In the third and fourth column of Table 2, we allow the average treatment effect to differ for stores that lag far behind and for stores that can be expected to be close to the winning positions. Specifically, we interact the treatment dummy with $I_{t, w - 1}$, the dummy indicating that a treatment store’s expected relative performance $\bar{D}_{t, w - 1}$.

In Table 3, we run regressions where we allow the treatment effect to depend linearly on expected intermediate relative performance, $\bar{D}_{t, w - 1}$.

The relation between the expected difference $E\bar{D}_{t, w - 1} - \max_{i} p_{CU t, w - 1}$ and the residuals from estimating (3).
estimation in the first column of Table 3 shows that intermediate relative performance is significantly positively related to subsequent sales. Its point estimate suggests that a percentage point increase in lagged relative performance increases current sales of treatment stores by 0.26%. However, in the IV-2SLS estimation, reported in the third column, the point estimate is more than halved and is no longer significantly different from zero. This underlines the importance of correcting for serial correlation in sales. Fig. 8 visualizes these results for the relevant subset of observations: treatment stores in the three final weeks of the experiment. It plots the residuals of the estimation of the average treatment effect (Eq. (3)), as presented in the first column of Table 2, against the predicted difference \( \frac{p_{i,t}^{U} - p_{i,t-1}^{U}}{\max_{i,c}} \) as estimated by the first-stage regression of the IV-2SLS estimation (second column of Table 3). In line with the estimation results, there is no easily discernible relation between predicted performance and sales, corrected for week and store-fixed effects.

The estimation in the third column of Table 3 assumes that the effect of intermediate relative performance is the same for all experimental weeks. In the fourth column of Table 3, we allow the effect to differ between weeks. The results show that there are no statistically significant effects of intermediate relative performance after the first two experimental weeks, while the effect at the start of the final week of the experiment is positive and significant at the 10% level. Moreover, our estimates show that stores that perform as well as their best-performing comparison store in the first three weeks of the experiment increase sales in the final experimental week by 4.5%. Stores that outperform their best comparison store further increase sales by 0.25% per percentage point distance to their best competitor, while stores that lag behind their best comparison store likewise show a weaker increase in sales. These regression results show that the effects of the treatment and of intermediate relative performance are concentrated in the final week.

The estimations in Table 3 assume that the effect of intermediate relative performance is linear. The first column of Table 4 reports the results of estimating Eq. (8), where the effect of intermediate relative performance at any level of intermediate relative performance is at most 5 percentage points below its best comparison performing stores. The second column of Table 4 shows that the marginal effect of incentives separate for stores close to winning a bonus. Dynamic incentives separate for stores close to winning.

The second column of Table 4 shows that the marginal effect of intermediate relative performance on current sales of relatively good-performing stores is significantly positive in all experimental weeks, with magnitudes ranging from 0.55 to 2.19% per percentage point. This relation is particularly strong in the third and fourth week of the experiment. For stores that lag far behind, there is no such effect in any of the weeks. This is not due to a lack of statistical power. Given the size of the standard errors, we should be able to detect at least some statistically significant effects if stores that lag far behind respond in a similar way as stores close to winning.

### Table 4

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Relative intermediate performance</th>
<th>Treatment + close</th>
<th>Relative intermediate performance + close</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment week 1</td>
<td>0.009 (0.017)</td>
<td>0.0001 (0.001)</td>
<td>0.0049 (0.025)</td>
</tr>
<tr>
<td>Treatment week 2</td>
<td>0.0048 (0.034)</td>
<td>0.0002 (0.001)</td>
<td>0.0049 (0.025)</td>
</tr>
<tr>
<td>Treatment week 3</td>
<td>0.0074*** (0.0017)</td>
<td>0.0007* (0.0009)</td>
<td>0.0063** (0.0032)</td>
</tr>
<tr>
<td>Treatment week 4</td>
<td>0.0045 (0.003)</td>
<td>0.0002 (0.001)</td>
<td>0.0087*** (0.003)</td>
</tr>
<tr>
<td>Relative performance after week 1</td>
<td>0.0012 (0.0019)</td>
<td>0.0022 (0.0021)</td>
<td>0.0021** (0.0011)</td>
</tr>
<tr>
<td>Relative performance after week 2</td>
<td>0.0034 (0.0034)</td>
<td>0.0019 (0.0021)</td>
<td>0.0005* (0.0003)</td>
</tr>
<tr>
<td>Relative performance after week 3</td>
<td>0.0022 (0.0045)</td>
<td>0.0019 (0.0021)</td>
<td>0.0021** (0.0011)</td>
</tr>
</tbody>
</table>

Standard errors clustered at the store level in parentheses.

\(*\) Close is a dummy variable that takes value one when the store’s expected intermediate relative performance is at most 5 percentage points below its best comparison store, i.e. when \(D_{w,1} < -5\%\) (see Eq. (6)). \(**\), \(***\) denote statistically significant effects at the 1%, 5%, and 10% level, respectively.

4.2. Robustness checks

Our results are qualitatively robust to varying the level of intermediate relative performance at which stores are deemed to be close to winning.
winning between −2% and −10%. Table 5 shows the estimated effects when the cut-off level is −5%, −2%, −7.5%, and −10%, respectively. In all cases, the effect of intermediate relative performance on current sales for stores deemed to stand a chance is positive and statistically significant. Quantitatively, the estimated effects are larger when the cut-off level is closer to −2%. The effects by week are less robust, and not statistically significant for all weeks. This is perhaps unsurprising, given that an increase in the cut-off level (to −2%) reduces the number of relevant observations and hence statistical power, while lowering the cut-off level reduces stores’ perceived winning probabilities, in particular in the last week of the experiment.

Whether store employees perceive themselves as close to winning, however, may depend not only on the distance to the best performing comparison store, but also on the distance to the second-best comparison store. As noted earlier, we cannot disentangle the two effects due to the group assignment procedure ensures that treatment and control stores are located in different regions, bias only arises if region-specific shocks are correlated across regions, as well as over time.

To address this issue, we include region-specific week fixed effects in Eqs. (5) and (8). That is, we estimate \( \eta_{r} \) in Eqs. (5) and (8) for each region separately. We distinguish between 12 geographically-clustered regions, using the same region classification as the company does. The estimation results are reported in the fifth and sixth columns of Table 3 and in the third column of Table 4, respectively. Our main results are by and large robust. The point estimate of the effect of intermediate relative performance for stores close to winning, reported in the third column of Table 4, is in the same order of magnitude (0.63% rather than 0.74%), and still statistically significant (at the 5% level). However, the estimated effects of treatment and intermediate relative performance in the final week of the experiment, reported in the sixth column of Table 3, are much smaller and no longer statistically significant. Possibly, we do not have a sufficiently large number of observations to identify region-specific week fixed effects as well as the effect of intermediate relative performance split out by week.

Next, to assess the exogeneity of our instrument we conducted a placebo-experiment. In particular, we pretend that our experiment would have taken place in weeks 1–4 instead of weeks 5–8 of 2010, following similar estimation procedures as above. If our instrument is free from serial correlation, we expect no significant effects of region-specific shocks on sales for stores deemed to stand a chance, i.e. when \( D_{t,w} = 1, \eta_{r} > 0 \) (see Eq. (6)).

A causal interpretation of the estimations discussed above crucially depends on the exogeneity of our instrument. Our instrument is constructed under the assumption that performance of the comparison stores is unrelated to subsequent performance of the treatment stores, except for common shocks and the effect we aim to identify. More formally, we assume that \( \varepsilon_{t,w} \perp \varepsilon_{t,w+1} \) and \( \varepsilon_{t,w} \perp \varepsilon_{t,w+2} \). One possible channel via which performance of the best comparison stores can influence subsequent performance of treatment stores is the existence of regional shocks that are correlated across regions and over time. In that case, regional shocks may influence the predicted difference, \( D_{t,w} \).
intermediate relative performance on subsequent performance. Serial correlation can bias our IV-estimates, as treatment stores’ expected performance \( E[p_{\text{CU}} - 1] \) is based on past performance. Although we exclude the 3 weeks prior to the experiment from \( E[p_{\text{CU}} - 1] \), serial correlation in stores’ own performance may still bias our instrument if there are long term trends in stores’ performance. The results are reported in Table 6. The first column reports the OLS estimates, the second column the linear IV-estimates. As expected, the OLS estimates suggest a significant positive effect of intermediate relative performance on subsequent performance, but this effect vanishes in the IV-2SLS estimation. The third column of Table 6 estimates the effect of relative intermediate performance for stores that are close to winning. We find a positive treatment effect for stores close to winning, and this effect is decreasing in intermediate relative performance. Both effects are statistically significant at the 10% level. At first sight, this seems to suggest that our instrument is biased by serial correlation, although the effect of intermediate relative performance for stores close to winning goes in the opposite direction of the main estimations. However, in contrast to the findings reported above, this effect becomes insignificant when we include region-specific week fixed effects: the point estimate drops to zero, while the standard errors remain of similar size (see column 4). Thus, the correlation is completely driven by region-specific shocks that are correlated both across regions and over time. By contrast, if serial correlation in stores’ own performance would bias our instrument, adding regional-specific shocks should not matter.

4.3. Discussion

Taken together, our results paint the following picture. We have found that stores lagging too far behind do not respond to the incentive scheme, nor to the intermediate relative performance information. However, as stores are closer to winning a bonus, sales increase significantly with lagged relative performance. This effect is strongest in the second half of the experiment. This result contrasts with Casas-Arce and Martinez-Jerez (2009), who do not find that performance decreases when trailing contestants lag further behind. One explanation for their result, as conjectured by the authors, is attrition bias, which is absent in our study. Frank and Obloj (2011) also study a competition without attrition and find, like us, that performance is increasing in intermediate relative performance for contestants that lag behind.

On average, the relative performance incentive scheme that we study had no effect on sales. Possibly, many stores have perceived the relative performance targets as too ambitious from the start, particularly in the light of their limited means to boost sales. Such a perception would be reinforced after receiving the first poster with rankings, as only 23 treatment stores ranked on top of the first poster, and 64 stores lagged more than 5% behind their best-performing comparison store on the first poster. An alternative explanation for the weak response is that the prospect of competing against non-competitors did not excite employees in the treatment stores as much as a real tournament would. It is hard to distinguish between these two interpretations. Attainability of the target clearly plays a role, as stores close to the winning positions respond positively to the treatment. This finding, however, does not rule out that a competition against competing competitors would have induced more substantial treatment effects. One specific reason why this would be the case is that feelings of envy play a more prominent role in a conventional tournament than in a comparable benchmark competition. In a conventional tournament, prize money is divided unequally among tournament participants, which gives losers reason to envy the winners. This stimulates effort among competitors, because by exerting effort, competitors can reduce the probability of experiencing envy (see Grund and Sliwka, 2005; Kräkel, 2008; Ederer and Patacconi, 2010; Gill and Stone, 2010, and Bartling, 2011). The same reasoning does not apply to our benchmark competition, as employees in the treatment stores do not have a reason to envy employees in the comparison stores (since the latter cannot win a prize). In a recent lab experiment, Eisenkopf and Teyssier (2013) examine the role of envy in competitions by comparing stated effort in a conventional two-person tournament with stated effort in a similar benchmark competition. In the benchmark competition, effort is evaluated against the effort of a randomly chosen participant from the conventional tournament treatment. Hence, the two competitions are identical, except for the externality effort imposes on others. They find that the conventional tournament induces individuals to put in significantly more effort than the

Table 6
Results of a placebo-experiment in weeks 1–4, 2010.

<table>
<thead>
<tr>
<th>Dependent variable: ( \ln(sales) )</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>0.015</td>
<td>0.0011</td>
<td>−0.0052</td>
<td>−0.022</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.014)</td>
<td>(0.017)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>Placebo relative int. performance</td>
<td>0.0017**</td>
<td>−0.0003</td>
<td>−0.0004</td>
<td>−0.0009</td>
</tr>
<tr>
<td></td>
<td>(0.0007)</td>
<td>(0.0007)</td>
<td>(0.001)</td>
<td>(0.0008)</td>
</tr>
<tr>
<td>Placebo treatment + close</td>
<td>0.063**</td>
<td>0.031</td>
<td>0.044</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0038)</td>
<td>(0.0025)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo relative int. performance*close</td>
<td>−0.0072*</td>
<td>−0.0011</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0038)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Store-fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Week-fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Region-week-fixed effects</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Store–week observations</td>
<td>3402</td>
<td>3402</td>
<td>3402</td>
<td>3402</td>
</tr>
<tr>
<td>Stores</td>
<td>189</td>
<td>189</td>
<td>189</td>
<td>189</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.9246</td>
<td>0.9243</td>
<td>0.9233</td>
<td>0.9535</td>
</tr>
</tbody>
</table>

Standard errors clustered at the store level in parentheses.

*Close is a dummy variable that takes value one when the store’s expected intermediate performance is at most 5 percentage points below its best comparison store, akin to \( D_{\text{CU}} - 1 > -0.05 \) (see Eq. (6)), but adjusted to the placebo period.

***, **, * denote statistically significant effects at the 1%, 5%, and 10% level, respectively.

28 There is substantial persistence in the weekly rankings, but stores that do not rank first after the first week still have a reasonable probability of winning. Specifically, about 40% of the eventual winners of a prize did not rank first after the first week. These stores lagged behind by 8 to 9 percentage points, on average. Likewise, stores that rank first after the first week, have 52% probability of finishing in first position. So, there is sufficient persistence in the rankings for intermediate performance feedback to be valuable, but changes in the final rankings are still possible.
benchmark competition, suggesting that (anticipated) feelings of envy boost effort in a tournament. This may also explain why we fail to find a statistically significant treatment effect in our setting. 

5. Concluding remarks

We have reported the results of a field experiment on dynamic incentive effects of relative performance pay among stores of a retail chain. We find that intermediate relative performance feedback affects subsequent performance of stores closer to the bonus target. These stores show significantly higher performance, particularly near the end of the incentive period. Stores lagging far behind do not respond to the incentive scheme, nor to intermediate relative performance. As many treatment stores happen to trail far behind bonus targets over the course of the experiment, we find no improvement in performance on average.

Our findings underline the importance of dynamic incentive effects. When in the course of a contest the target moves out of reach, people give up, which renders the incentive scheme fruitless. On the other hand, learning that intermediate performance is closer to target encourages people to increase effort. Hence, the incentive effect of competition is path-dependent.

References


Frank, Douglas, and Tomasz Oloj (2011), Reference points and organizational performance: evidence from retail banking, mimeo, INSEAD.


Hochtl, Wolfgang, Kerschbamer, Rudolf, Stracke, Rudi, Sunde, Uwe, 2011. Incentives vs. selection in promotion tournaments: can a designer kill two birds with one stone?, IZA discussion paper no 5755.


