NBER WORKING PAPER SERIES

KINKS AS GOALS: ACCELERATING COMMISSIONS AND THE PERFORMANCE OF SALES TEAMS

Peter J. Kuhn
Lizi Yu

Working Paper 28487
http://www.nber.org/papers/w28487

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2021, Revised June 2023

We thank Youssef Benzarti, Clement de Chaisemartin, Takao Kato, Alex Rees-Jones, Kathryn Shaw, Alisa Tazhitdinova, Gonzalo Vasquez, Guangli Zhang, and participants at the 2019 Trans-Pacific Labor Seminar (Tokyo) for helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Kinks as Goals: Accelerating Commissions and the Performance of Sales Teams
Peter J. Kuhn and Lizi Yu
NBER Working Paper No. 28487
February 2021, Revised June 2023
JEL No. J33,M52

ABSTRACT

We study the performance of small retail sales teams facing an incentive scheme that includes both a lump sum bonus and multiple accelerators (kinks where the commission rate jumps upward). Consistent with standard labor supply models, we find that the presence of an attainable bonus or kink on a work-day raises mean sales, and that sales are highly bunched at the bonus; inconsistent with those models we find that teams bunch at the kinks instead of avoiding them. Combining simple theoretical models, institutional evidence, and heterogeneity analyses, we argue that that this unexpected bunching results from a previously unrecognized motivational benefit of piecewise linear reward schemes in team environments: Teams use the convenient, salient, kink-points as shared goals, which yield symbolic utility to their members when the points are attained.

Peter J. Kuhn
Department of Economics
University of California, Santa Barbara
2127 North Hall
Santa Barbara, CA 93106
and IZA
and also NBER
pjkuhn@econ.ucsb.edu

Lizi Yu
University of Queensland
lizi.yu@uq.edu.au

1 Introduction

Putting an accelerator or accelerators into the commission plan is the most common method applied. Put simply, the rate of commission paid is accelerated for every dollar earned above the set sales target. This encourages the salesperson to not only meet their sales targets, but to 'smash their sales targets'! (MyFirstSalesJob, 2020)

Despite the well known advantages of linear reward schemes (Holmstrom and Milgrom, 1987), non-linear schedules are frequently used to incentivize workers. For example, bonuses –which give workers a lump sum reward for attaining a threshold level of output– are used by a majority of firms to incentivize their salespeople, and have been extensively studied.[1] Studies of bonuses have documented both their incentive-enhancing effects and the distortions associated with their highly non-linear nature (Freeman et al., 2019; Owan et al., 2015; Benson, 2015).

Perhaps surprisingly, another common form of nonlinear incentives –accelerators– has received much less theoretical and empirical attention.[2] Accelerators are performance thresholds at which the *slope* (but not the level) of the pay schedule jumps upward. In principle, accelerators could affect workers' effort and performance through at least two channels. First, standard labor supply models predict that an accelerator at an output threshold will encourage workers to exceed the threshold, but to *avoid* output levels just around the threshold. Second, accelerators (like bonuses) divide the set of feasible performance levels into salient, disjoint ranges, which frequently have labels attached to them, such as "Tier1" and "Tier 2", or "Good", "Excellent", and "Superior". These ranges could function as reference points or symbolic rewards that affect worker behavior. As we show in the paper, accelerators provide a unique opportunity for distinguishing between these two channels, because –in contrast to bonuses where the two channels have similar predictions– the channels' predictions for the distribution of output conflict.

This paper studies the effect of accelerating reward schemes on worker performance from both an empirical and theoretical perspective. The setting for our study is 103 small retail stores operated by a large clothing firm in China (hereafter "Firm A") whose sales teams face a monthly pay schedule that includes three thresholds. At the lowest of these (henceforth the *target*) the commission rate rises and a small lump-sum bonus (worth about 3.5 percent of mean monthly pay) is awarded. At the other two thresholds, (henceforth the *kinks*) only the commission rate rises. The four team output intervals corresponding to these three thresholds are labeled as "Standard", "Good", "Excellent", and "Superb" on a worker's monthly pay stub. As we detail in the paper,

[1] Douthit (1976) and Joseph and Kalwani (1998) respectively report that 80 and 72 percent of firms use bonuses to incentivize their salespeople; Oyer (2000) reports that 89 percent use sales quotas.

[2] O'Connell (1989) states that 53 percent of sales compensation plans have increasing commission rates, and Parrinello (2017) reports that over-quota accelerators are used in most sales compensation plans. Lazear (2000) and Larkin (2014) have studied accelerating reward schemes. Noteworthy complementarities between our paper and Larkin (2014) include (a) the fact that we document and attempt to explain a new and unexpected effect of accelerators (bunching), (b) our focus on a context where time shifting of sales is essentially impossible (so our unexpected results most likely reflect sales effort choices), and (c) the team-based nature of incentives in the firm we study.

no other monetary rewards, such as promotion or retention decisions, are linked to attaining these thresholds. Nevertheless, workers may attach a symbolic value to attaining these sales thresholds as a team.

We start our analysis by deriving the implications of the standard labor supply model for the distribution of team output around Firm A's two types of thresholds, i.e. the *target* and the *kinks*.[3] We then compare these predictions with empirical estimates of output distributions at both the monthly and daily levels. In the latter case, our estimates use quasi-random variation in the presence and location of a consequential threshold that is driven by two factors: whether a day is the last of a month, and the team's inherited cumulative sales on the morning of the last day. We then use a *difference in density differences* (DDenD) estimator to estimate the effects of thresholds on the entire frequency distribution of daily sales.

Consistent with the standard labor supply model, we find strong bunching of output levels at the target (where a bonus is paid). Also consistent, we find that the presence of an attainable target *or* kink on the final day of a month raises mean sales on that day. Inconsistent with the standard model –which predicts missing density in the sales distribution at 'concave' kinks– we find bunching at the two pure kinks.[4] This bunching is robust to controls for a rich set of cyclical effects, and to choices of estimation samples, control groups, bin widths, and statistical inference methods. Finally, we use heterogeneity analyses, detailed institutional information, and estimates of the how attainable kinks shift the distribution of mass over the entire sales distribution to identify the most likely causes of this unexpected bunching. After considering several candidate explanations, we argue that the mostly likely cause is rooted in team production: Firm A's teams use all three thresholds as convenient, salient collective goals to coordinate and motivate their members. In consequence, the teams behave as if a lump-sum psychological reward was attached to attaining each kink.

In addition to the literature on non-linear reward schedules, our paper contributes to the literatures on symbolic rewards and goals. While a number of authors have studied symbolic rewards, much of the existing work focuses on *relative* symbolic rewards that induce a tournament among employees (Kosfeld and Neckermann, 2011; Ashraf et al., 2014; Ager et al., 2022). Awards of this type can have negative spillover effects on the workers who do not receive them, which can wash out their incentivizing effects on the winners. In our case, the symbol is activated when the team (comprising all the employees of the company at a particular workplace) attains a pre-specified *absolute* level of performance. Since everyone in a worker's natural reference group either receives or does not receive the symbolic reward, our rewards should induce co-operation, not competition.

[3]We start our analysis by modeling the behavior of sales teams as if they were individual workers. Section 3.3 describes two alternative ways in which our models of individual behavior might translate into team-level behavior and proposes a simple test between them, based on heterogeneity by team size.

[4]Kleven (2016) distinguishes between convex kinks (which maintain the convexity of the budget set and induce bunching) and non-convex kinks, which should produce gaps. We use the term *concave* for the second type of kink. Kleven reports that existing research on concave kinks is rare, and has not found much evidence of missing mass there. Seibold (2021) is the only study we know of that, like us, finds bunching at concave kinks. Those kinks appear to be driven more by a lack of information about Germany's retirement system than by the symbolic incentives we argue are at work in our context.

Thus, an additional finding of our paper is that *team-based* symbolic rewards based on *absolute* performance can be motivating, without the negative spillover effects associated with awards for relative, individual performance. Similar results have been found in the literatures on gainsharing and high-performance work systems, which argue that giving semi-autonomous groups of workers collective goals can be highly effective (e.g. Knez and Simester (2001)).[5]

The literature on *goals* studies situations like ours where the set of an agent's feasible performance levels is divided into intervals with potentially meaningful labels attached to them (Genicot and Ray, 2020; Koch and Nafziger, 2020). In our reading, the existing empirical work focuses mostly on individual performance targets that are chosen by agents themselves. Often, these *commitment contracts* impose a monetary cost on agents who do not attain them (Royer et al., 2015; Kaur et al., 2015).[6] More closely related to our work, Agarwal et al. (2017) study exogenously-assigned goals for a prosocial activity–resource conservation– with no financial consequences.[7] Like us, they find that goals work best when they are neither too easy nor too hard, but these authors only consider loss aversion as a possible explanation for the goals' effects.

Our study also connects to a literature on piecewise-linear budget sets in other contexts, including retail sales and income taxes. Many of these papers focus on convex kinks, such as progressive income taxes and rising marginal prices for products like electricity and cell phone service (Lambrecht et al., 2007); these correspond to *decelerators* in incentive pay schemes, and the standard model predicts bunching at these kinks. Studies of concave kinks like our accelerators appear to be rare, however, and often have special features that rule out tests for the missing masses predicted by the standard model.[8] Thus our context appears to offer a rare opportunity to study the 'missing mass' predictions of the standared model for concave kinks in budget sets.[9]

Finally, our paper contributes to a growing applied literature devoted to estimating the amount of bunching, or 'excess density' in distributions of economic outcomes at theoretically relevant values. A number of papers in the tax literature (Chetty et al., 2011; Saez, 2010; Kleven and Waseem, 2013; Mortenson and Whitten, 2019) address this question using large samples of draws from what is assumed to be a single distribution, in which there is a fixed potential bunching point of interest. For example, one may have millions of observations on the level of pre-tax income, and one is interested in the extent to which taxpayers bunch at the maximum EITC entitlement. The authors then fit a high-dimensional polynomial to the entire frequency distribution –excluding a small region of interest– and use the implied smoothness assumptions to identify the excess mass

[5]Related, Larkin et al. (2012) have argued that team-based pay can reduce the psychological costs from social comparison associated with individual performance-based pay.

[6]Dobronyi et al. (2019) study such goals when no monetary cost is attached, and find they have a zero effect.

[7]Our context is also similar to the conversion of numeric to letter-based academic grades (Oettinger, 2002), though letter grades have real 'lump sum' consequences for academic promotion. In addition, letter grading schemes frequently have a competitive, zero-sum feature that is absent in our context.

[8]For example, the quantity discounts that are associated with different package sizes in retail sales do not allow for bunching-based tests because there is a small and fixed number of package sizes (Cohen, 2008).

[9]A few papers on accelerating reward schemes have documented their effects on *risk-taking* (Chevalier and Ellison, 1997; Shue and Townsend, 2017; de Figueiredo Jr et al., 2023). We do not study these consequences of accelerators here, focusing instead on documenting a novel, effort-enhancing effect of these pay schemes.

in that region.

In our case, while we observe team sales on a large number of days, our models suggest that we should not see any bunching or gaps on most of those days (either because there is no commission threshold within reach, or because reaching the threshold on that day is not a necessary condition for attaining it for the month).[10] This gives us a large sample of *control days*. On the other days (which are a subset of the last days of the month), we expect bunching or gaps at particular sales levels that vary quasi-randomly from day to day, and from team to team. Our setting is complicated by the fact that other observable factors, such as day of week and store location, also affect the daily sales distribution and should be controlled for. As noted, we address this issue using a "difference in density differences" (DDenD) estimation approach that allows us to non-parametrically estimate the effects of an attainable threshold on the entire frequency distribution of daily team sales, with rich controls that include unrestricted store-specific baseline sales distributions.[11] While our approach is not optimal for all contexts, it allows us, for example, to measure the extent to which the presence of an attainable kink in the $1100-$1199 bin on a given day and store shifts the entire distribution of output, not just in the area around the bin itself. These features may prove useful in other studies of the effects of nonlinear incentive plans, including bonuses, reference points, and payday effects, among others.

2 Background and Data

2.1 Retail Jobs and Firm A

Our data represent 103 retail stores operated by Firm A, a large manufacturer and retailer of men's clothing in China.[12] We study the performance of these stores during 2016, when the incentive scheme we analyze was in place.[13] Dropping observations in the month of store openings, closures or remodeling, we have 34,863 daily observations of team-level sales. The company does not collect individual employee sales data, and all commission payments are based on team-level sales.

Firm A operates retail stores in two types of locations: department stores and shopping malls (henceforth *host institutions*). A typical store includes a counter and a display of products, and

[10]In our main estimates, we define a sales threshold as *within reach* on a given day if it is less than three typical days' worth of sales away. Our results are robust to other definitions, and respond to the distance to the threshold (i.e. 1, 2 or 3 days) in intuitive ways (see Appendix C.1 and Table 6.)

[11]Pierce et al. (2022) use a similar approach to estimate the effects of loss-framed performance incentives on sales distributions; unlike us, however they apply kernel smoothing to their density estimates. Our approach is also closely related to Cengiz et al. (2019), who directly estimate frequencies in all the bins of a wage distribution when a theoretically relevant quantity –in their case the minimum wage– falls into different bins in different times and places.

[12]Our sample excludes 46 stores whose workers were not directly hired and paid by Firm A. Since the teams' monthly targets, T, were given by their own sales in the previous year (2015), our sample also excludes stores that did not operate in 2015.

[13]Firm A's salespeople had experience with similar types of incentive schemes in previous years, but the details varied somewhat from year to year.

is operated by a sales team of 2-7 employees. One member of each sales team is designated as the store manager; like other team members, managers spend most of their time doing front-line sales, but are also responsible for coordinating schedules and encouraging cooperation among the team members. According to the frontline salespeople, the yield rate per customer, i.e. the fraction of people who walk into the store making a purchase, is quite high for men's clothing retail, at between 50 and 90 percent. Despite this high yield rate, employees and managers believe that sales amounts are quite sensitive to employee effort. In part this is because Firm A's salespeople play a more active role than many U.S. salespeople: With very limited display space, most items must be retrieved from storage to be tried on. Customers also place a high value on fit and speed of service, both of which are sensitive to employee effort.

Since retail sales also depend on customer traffic, daily sales fluctuate quite dramatically over time, as shown in Figure 1(a). This figure plots the average daily sales per store on every calendar day in 2015 and 2016. The smaller cycles in the figure represent output within a week, with higher daily sales on weekends. The largest spikes are labelled, and correspond to major holidays when people shop for menswear heavily, such as Chinese New Year or Father's Day. Controlling for these large day-of-week and holiday effects will play a significant role in our econometric analysis. As shown in Table 1, the average daily sales output was \$582 in 2016, but daily sales were both 'lumpy' and unpredictable: only 11.3 relatively high-value items were sold per day, and the standard deviation of daily sales (\$1241) was over twice the mean and 4.5 times the median. In Figure 1(b), we aggregate the daily sales data by calendar month. Between March and December, the monthly sales pattern is essentially identical in 2015 and 2016, indicating strong seasonal effects. In January and February, however, the two years' patterns are starkly different, because the movement of the Chinese New Year shifted sales from February 2015 into January 2016. As we demonstrate in Appendix C.2.4, the 'target mismatch' induced by this holiday shift contributes to our identifying variation, but is not essential to it.

2.2 The Compensation Schedule

The employees in our sample are paid monthly, including a base salary and a commission. In 2016, the average monthly compensation was \$545, around 40% of which was commission. Within a store, the base salary varies with employees' tenure, while commission rates are identical for all employees. Commissions are based on the *store's* total monthly performance; they are set lower in larger stores in order to yield similar daily wages across teams of different sizes, as indicated in Table 1. Managers are paid the same as the other team members, with a small increment in base pay for handling their managerial duties.

The solid black line in Figure 2 illustrates Firm A's pay scheme in 2016. It included a monthly sales target, T, at which a bonus is awarded and the commission rate increases, plus two pure accelerators where only the commission rate rises. If a store attains T, each team member receives a lump sum bonus of \$15.63 (100 CNY), and is exempt from a \$3.13 (20 CNY) penalty for not meeting the target, resulting in a jump equivalent to \$18.76 (120 CNY). This is worth 3.5% of employees' average monthly compensation. At the target T, the commission rate accelerates to 1.5 times the baseline rate. Next, there are two pure kinks in the schedule: The first accelerator is at $1.3T$, where the commission rate rises to 1.8 times the baseline rate, and the second is at $1.6T$, beyond which twice the baseline commission rate is paid.[14]

In 2016, a store's target T was equal to its sales in the same calendar month in 2015. While this might raise a concern that teams' 2016 targets were affected by strategic output restriction in 2015 (Charness et al., 2011), this concern does not apply to our setting because 2016 was the first year in which Firm A tied the sales targets to a store's previous year's performance. Importantly, Firm A's decision to link the 2016 targets to 2015 sales was announced after all the 2015 sales were realized.[15]

At the beginning of each month, the sales workers are informed of their team's monthly sales goal, by their team leader. (The team leader receives this information from head office.) The team leaders can calculate their team's cumulative sales anytime during the month, but typically only do this on a few days, such as the end of first week, the middle of the month, the start of the last week, or the second-last day. It is up to the team leaders whether, when, or how to drum up enthusiasm to work towards one or more of the thresholds. Our conversations with workers and management suggest that all three thresholds are frequently promoted by team leaders as collective goals, when they are within a reasonable 'striking distance'.

In interviews, management described three main motivations for their use of this pay scheme. First, they believe that the target T is a simple and effective way to communicate their expected performance level to each store, in a context where customer traffic varies dramatically across stores and across months of the year (Ockenfels et al., 2015). Second, as argued in Kuhn and Yu (2021), Firm A's use of *team*-based incentives is part of an HR system that delegates many key decisions, including employee discipline and some aspects of recruiting, to the team members.[16] Third, (when asked) Firm A attributes its use of accelerating piece rates (without bonuses) at higher performance levels ($1.3T$ and $1.6T$) to a stated desire to be consistent with other local retailers and to retain high-performing salespeople. Indeed, Firm A's commission schedule –a target with a bonus attached, plus accelerators beyond that– is common among its competitors in the region. Thus we are studying a pay system that has withstood the test of time and competition, not an

[14]By 2016, Firm A had used a similar nonlinear schedule –containing bonus(es) and accelerator(s)– for at least five years, but the targets were either determined by store employees themselves (subject to management approval), or by regional or higher-level managers.

[15]Even if workers in 2015 somehow misperceived that their efforts could affect next year's targets, the incentives for strategic output restriction would be very low, due to Firm A's high turnover rate (Kuhn and Yu, 2021).

[16]Despite the theoretical possibility of free-rider problems, team-based incentives are used quite widely, including Continental Airlines (Knez and Simester, 2001), Microchip (Adamson et al., 2014), U.S. steel minimills (Boning et al., 2007), and apparel manufacturing (Berg et al., 1996; Hamilton et al., 2003). Examples from retail include Wal-Mart's profit-sharing plan, team-based bonuses in German retail establishments (Friebel et al., 2017), and tip pooling in restaurants (Scudder, 2017). Lawler and Mohrman (2003) report that the share of Fortune 1000 companies using work-group or team incentives for more than a fifth of their workers more than doubled, from 21 to 51 percent, between 1990 and 2002.

unfamiliar, experimental policy adopted by a single firm.

According to Firm A, except for the bonus at T, no additional monetary rewards are attached to attaining the thresholds in its nonlinear incentive plan. As we document in Section 6.4, threshold attainment is also irrelevant to employee retention and promotion decisions; thus it will be hard to ascribe any bunching that we observe at sales thresholds to unobserved, discrete *material* rewards that might be attached to those thresholds. Workers could, however, be motivated by *symbolic* rewards associated with attaining a threshold contained in their monthly pay slips. Figure 3 presents a template of the pay slip distributed to salespeople. Specifically, the commissions are calculated and recorded separately for the four sales ranges that correspond to the available commission rates (i.e. $1\times$, $1.5\times$, $1.8\times$, or $2\times$ the baseline). Thus, the worker sees her *team's* continuous performance measure –which is essential for calculating her pay– and in addition sees which of the four ranges the team achieved that month, which are labelled "Standard", "Good", "Excellent", and "Superb", respectively.[17]

3 Theory– Predicted Output Distributions

In this section we describe the predictions of some simple theoretical models for the empirical distributions of output around the two types of thresholds in Firm A's pay schedule– the target (T) and the kinks ($1.3T$ and $1.6T$). We begin by describing the predictions of a standard labor supply model that treats Firm A's sales teams as though they were individual workers who derive utility only from money and leisure.[18] Next, we extend this model to allow for reference points and psychic rewards. Specifically, in our *loss-aversion* (LA) model, the three commission thresholds act as reference points; in our *symbolic rewards* (SR) model workers experience a lump-sum psychic reward when their monthly sales exceed one of the commission thresholds.[19]

Third, we consider two different ways in which loss aversion or symbolic rewards experienced by *individual* workers could translate into the behavior of Firm A's small, multi-worker teams: *uncoordinated maximization*, or coordination that is facilitated by *shared team goals*. We conclude with a brief preview of our evidence, and argue that the evidence is most consistent with *shared team goals* that are experienced as *symbolic rewards* by individual workers when a team sales threshold is attained.

3.1 The Standard Model

We begin our analysis of the standard model by focusing on the two 'pure' kinks ($1.3T$ and $1.6T$). As illustrated in Figure 2, the standard model predicts gaps, or missing density, in the

[17] We do not know precisely when this practice started, but salespeople recall instances from as early as 2014.

[18] To keep the model as simple as possible, we assume the teams' utility is linear in money. We also abstract from production uncertainty by assuming that workers have complete control over their sales output.

[19] Following Saez (2010) all our models assume that workers differ according to a single cost-of-effort parameter that is strictly positive and continuous on $(0, \infty)$. Formal definitions and analyses of the standard, LA, and SR models are provided in Appendix A. Our goal here is to present the main results and intuition as concisely as possible.

distribution of output in intervals that (strictly) include each of the two kinks, in other words between c and d, and between e, and f. To see why, imagine a continuum of workers with progressively lower effort costs (and hence flatter indifference curves); in this continuum, workers with lower costs will locate further to the right along the pay schedule. Around a kink like $1.3T$, there must therefore be a marginal worker who is just indifferent between performance (sales) levels c and d. As long as workers' indifference curves are continuously differentiable, no workers will locate between those two points; in other words there should be a gap in the distribution of output levels there. This result is well known in the literature on bunching estimators, although empirical studies of *concave* kinks (like ours) which are predicted to produce gaps, are rare (Kleven, 2016).

Continuing with the standard model, we turn next to its predictions around the target, T. Here, there must exist a worker who is indifferent between points a and b, indicated by the leftmost indifference curve in Figure 2. Workers with slightly lower effort costs, rather than locating slightly to the right of a, will instead choose to locate at point b, and the same is true for a non-empty interval of workers who would normally (i.e in the absence of a bonus and accelerator at T) choose to produce more than a but less than T. Thus we should see missing mass to the *left* of T, but *bunching* of the output distribution at T.[20]

3.2 The Symbolic Rewards (SR) and Loss Aversion (LA) Models

In Appendix A, we amend the preceding model in two different ways. First, in addition to money and leisure, our *symbolic rewards* model assumes that workers derive a lump sum of psychic utility from achieving an output level that meets or exceeds a salient threshold, like T, $1.3T$, or $1.6T$. Second, following an active behavioral economics literature that models the connection between loss aversion and bunching (Abeler et al., 2011; Agarwal et al., 2017; Allen et al., 2017; Pierce et al., 2022) we consider a *loss aversion* model in which salient thresholds function as reference points that induce loss aversion.

The predictions of these *symbolic rewards* and *loss-aversion* models are also summarized in Appendix A, separately for the pure kinks ($1.3T$ and $1.6T$) and the target, T. At the kinks, the symbolic rewards model now has the same qualitative predictions as the target in the standard model: missing density to the left of each kink, *bunching* at the kink, and excess density to the right. Intuitively, this is because a symbolic reward attached to reaching a threshold is equivalent to a cash bonus at that point. Finally, the predictions of the loss-aversion model at the kinks depend on whether the loss aversion is strong or weak relative to the change in the commission rate. If loss aversion is not strong enough to outweigh the effects of the rising cash incentives at the kink, the model's predictions for the kinks are the same as in the standard model– a gap in the output distribution around the kink. If loss aversion outweighs the acceleration in financial rewards, we should see bunching at the kink, and excess mass to its right. Finally, at the target, T, Appendix A

[20]We should also see some excess density to the right of the target, caused by the higher marginal incentives in that region.

shows that all three of our models –standard, symbolic rewards, and loss aversion– have the same qualitative predictions: missing mass below T, excess mass (i.e. bunching) at T, and excess mass to the right of T.

For easy reference, the above predictions are summarized in Table 2. In sum, our three models can only be distinguished by the output distributions we see at the pure kinks, $1.3T$ and $1.6T$. At those points, the standard model predicts missing density in a region that strictly includes each kink, and no bunching anywhere. The symbolic rewards (SR) model predicts bunching at the kinks (for any positive value of the symbolic reward), and the loss-aversion model (LA) *can* predict bunching at the kinks, but only if workers are sufficiently loss-averse. (The loss-aversion parameter must strictly exceed the change in the commission slope at the kink.)

3.3 Individual Goals versus Team Co-ordination

In this section we ask how the predictions of the preceding models of individual workers' behavior might translate into the behavior of small, multi-worker teams like those in Firm A. Focusing –for simplicity– on the only model that unambiguously predicts sales bunching at two pure kinks (symbolic rewards), we consider two competing ways in which symbolic rewards experienced by individual team members can affect team-level sales distributions.

The first possibility (the *uncoordinated maximization* hypothesis) assumes that each individual team member directly experiences a symbolic reward when their team passes a threshold in its commission schedule, but that team members select their effort levels independently. In this model, individual workers will have incentives to work towards 'just' attaining pure kinks, but those incentives will be rapidly attenuated by free-rider issues and team co-ordination problems as team size grows. Thus, under the uncoordinated maximization hypothesis, the amount of bunching at the kinks should diminish rapidly with team size.

In contrast, the *shared team goals* hypothesis assumes that teams (or their leaders) use the thresholds in piece-wise linear pay schemes as convenient, salient goals *in order to* solve co-ordination and moral hazard problems. In this interpretation, bunching would not occur in very small teams (such as one or two people) because such focal points are not needed: The cash incentives inherent in the commission schedule itself should be sufficient to elicit high individual effort. Larger teams, however, might need to adopt shared goals in order to solve the more severe moral hazard and coordination problems they face. According to this hypothesis, larger teams will collaborate to achieve a psychic reward that is enjoyed by all the members when team output exceeds an agreed-upon goal. In this case, bunching at the kinks would result.

3.4 A Preview of Our Results

As a preview of our empirical results, Figure 2 includes graphs of the empirically observed excess densities around T, $1.3T$ and $1.6T$ on the last days of calendar months. These are taken directly from Figure 5, which compares raw sales densities in last days of a calendar month in which a particular threshold (either T, $1.3T$, or $1.6T$) is within reach to non-last days in which the *same* threshold is within reach. The idea is that 'just' attaining a monthly threshold is decisive for both financial and psychic rewards *only* on the last days of a month: on all other days teams that fall short still have the opportunity of achieving that goal later in the month. Figure 5 clearly shows bunching at T, which is consistent with all three of our models. It also shows bunching at $1.3T$ and $1.6T$, which is consistent with the SR and LA models, but inconsistent with the standard model. In addition, our heterogeneity analysis in Section 6.3.2 shows that bunching is much more pronounced in larger than smaller teams. Accordingly, we shall argue that our teams' behavior is best described by a symbolic rewards model, where the symbolic rewards are created and used by teams to motivate their members.

4 Descriptive Evidence

As a first look at sales patterns in Firm A's stores, Figure 4(a) plots the distribution of sales across 1,143 store-month observations in 2016, where sales are measured relative to the team's target T (i.e. its same-month sales in 2015). For every store, observations are grouped into bins of width $0.1T$ that are aligned with the target ($0.8T$ to $0.9T$, $0.9T$ to T, etc); consistent with Firm A's target-setting formula, sales teams fail to reach their monthly target in 50% of all store-month observations.[21] Consistent with the standard model described in Section 3, we see clear bunching at T; this is, in fact where the bunching is most striking. Contrary to the standard model, however, we observe some bunching at $1.3T$. In contrast, there is no visual evidence of bunching, nor of the predicted missing density, at the last kink point $1.6T$. Aside from these two mass points, the histogram elsewhere exhibits a smooth right-skewed bell shape.[22]

To quantify these visual impressions, Figure 4(b) applies the approach developed by Saez (2010), Chetty et al. (2011), and Kleven and Waseem (2013) to estimate the amount of excess density around T, $1.3T$, and $1.6T$. Using the number of observations that fall within each bin, we fit a high-degree polynomial (8th degree), excluding data in those three bins.[23] Then, using the estimated coefficients from the polynomial, we extrapolate a fitted distribution to the excluded points to estimate the counterfactual distribution.[24] The estimated bunching using this approach is 0.3030 at T, which is significant at 1% using the bootstrapped standard error. This implies that the excess mass around the target is 30% of the average height of the counterfactual bin count. The

[21] In Appendix Figure C.3.1, we replicate this figure in half-size bins of width $0.05T$ and we find similar patterns.

[22] The observations in the tails are mostly generated by the fact that the company's targeting formula fails to account for holidays or sales promotions that occur in different months in different years. For example, the Chinese New Year arrived in February 2015 but in January 2016; as a result the teams' 2016 sales targets were very easy in January, and quite unrealistic in February.

[23] In a typical application of this method, a narrow range of bins near the hypothesised bunching location will also be excluded, to allow the excess mass to diffuse around these points. We do not exclude neighboring bins for two main reasons. First, we would have to exclude all bins from $0.9T$ to $1.7T$, as the three kinks in our context are close to each other, and that would lose a major part of the distribution. Second, the bins used in our analysis already allow for considerable noise around the kink, because each bin of $0.1T$ is equivalent to three average days' output.

[24] Details of this procedure are provided in Appendix B.

estimated bunching at $1.3T$ is 0.2194, which is also significant at 1%. In contrast, at $1.6T$, we find negative bunching of -0.1002, but the estimate is not significant.[25]

While Figure 4(b) provides preliminary evidence of bunching around the target and the first kink, our ability to make sharp statements using this approach is limited by the fact that we have only 1,143 observations of monthly store performance, compared for example to the 11.6 million and 4 million observations used by Chetty et al. (2011) and Kleven and Waseem (2013) respectively. These limitations motivate our focus on daily sales, which allow us to focus our bunching tests only on situations where there is a close connection between the current day's efforts and the probability of attaining a salient threshold in the current month. This connection is closest on the last days of months in which a threshold is less than three days' worth of sales away, i.e. *within reach*.[26] In monthly data, random negative shocks early in the month frequently make it impossible for teams to attain a threshold in some months; random early positive shocks can push teams beyond a threshold well before the month's final day.

To illustrate the relationship between teams' output thresholds and the *daily* distribution of sales, we start by measuring each team's 'inherited' output at the start of every day, t, as the sum of sales from the first day of the month up to and including day $t-1$. Then, separately for each of the three thresholds, we restrict our sample to days in which the threshold is within reach (i.e. within three days' worth of sales away), given the team's inherited output. Next, in Figures 5(a), (b) and (c), we plot the distribution of each day's *actual* output relative to the three thresholds (T, $1.3T$, and $1.6T$ respectively). We show these distributions separately for last days of a calendar month (with a threshold within reach) versus all the other days on which the same threshold is within reach.[27] Sales are measured in \$100 bins relative to each threshold, so a team that beat a threshold by \$350, for example, would fall in the middle of bin 3, and a team that beat the target by less than \$100 would fall into bin 0.

Again, and consistent with the standard model, Figure 5(a) shows a pronounced difference between the output distributions on the last day of the month –which is the team's last chance to reach the target, T– compared to all other days on which T is within reach. On the last day, a disproportionate share of teams sell just enough to beat the target by less than \$100 (p=.000). Contrary to the standard model, however, Figures 5(b) and (c) also show that teams bunch in the output bins that 'just' beat these thresholds on last days (which is the last chance to attain the threshold) relative to non-last days when the same threshold is within reach (p=.000 and p=.058 respectively). Together, these results suggest that teams who find themselves within reach of a salient threshold on the last day of the month tailor their efforts to just reach the threshold on that day. Notably, in Appendix C.4 we repeat this exercise for a wide range of placebo thresholds,

[25] This figure also suggests bunching at $2T$, consistent with preferences for round numbers as in Benson (2015) and Allen et al. (2017), but the excess mass at $2T$ is not statistically significant ($p = 0.857$).

[26] Table 6 considers the effect of alternative cutoffs for defining a threshold being within reach. Appendix C.1 uses an alternative definition to calculate if a threshold is within reach.

[27] Note that the daily sales distributions in Figure 5 tend to be centered a little below T. This is a consequence of our three-day window for a threshold being within reach: With such a window, the threshold is on average about 1.5 days of sales away on the morning of the current day.

ranging from $0.7T$ through $2.1T$. We do not find statistically significant excess density at any of these placebo thresholds.

5 Estimation Approach

Motivated by the descriptive evidence in Figures 4 and 5, our main econometric approach will compare a sales team's output on last days on which a threshold (T, $1.3T$ or $1.6T$) is within reach to the same team's output on all other days.[28] As noted, the distinction between these two types of days is whether or not there is a clear connection between a sales team's effort on that day and the team's chances of attaining a salient monthly sales threshold. Compared to the more familiar bunching methods discussed in the previous section, a key advantage of our approach is the large sample of control days, which allow us estimate the effect of the presence *and* location of a threshold on the entire daily sales distribution, without the need for smoothness assumptions. Instead, the sales distribution is represented by a finite number of bins, with an arbitrary baseline density attached to each of them. Similar approaches to density estimation have been taken by Cengiz et al. (2019) in the context of wage distributions, and by Pierce et al. (2022) in the context of sales distributions. Our focus on last days when a threshold is within reach follows Benson (2015), who creates a 'treatment bubble' of managers whose cumulative performance near the end of a pay period makes their current actions decisive for attaining a bonus.

To implement our approach, we create a set of \$100 daily sales bins for every store, which run from zero to the maximum daily sales we see for that store in our data period.[29] With 103 stores in our sample, this yields 34,863 daily store sales observations and 2,778,836 store$\times$day$\times$bin observations.[30]

5.1 Estimation without Covariates

To illustrate how our approach works, we first consider an example where the reward schedule has a single threshold (for example, at the sales target, T), and no adjustment is made for covariates– for example, factors like weekends that affect the distribution of sales. As a convenient shorthand, we will define a threshold as *attainable* on a given day if that threshold is within reach on that day, *and* that day is the last one in a month. In this case we would estimate the following linear

[28] Appendix C.2 shows that our results are robust to a variety of alternative control days.

[29] Since stores differ in size and performance, this yields a different number of bins for each store. While this might seem cumbersome, forcing every sales bin to be \$100 wide has the advantage that we can unambiguously interpret 'hitting' the bin that meets the team's monthly target as being within \$100 of the target. More generally, our 'binned' approach accommodates noise in the production process: By classifying any level of sales in the same \$100 bin as being 'at' the target, it includes teams that 'tried but just missed' as 'bunchers'.

[30] To assess the robustness of our approach to constructing sales bins, Appendix C.3.1 uses a bin size that is proportional to each store's target output, and shows that the main results are very similar.

probability model:

$$S_{it,b} = \alpha + \beta_0 \cdot K_0 + \sum_{j=-5}^{-1} \beta_j \cdot K_j + \sum_{j=1}^{5} \beta_j \cdot K_j + I_i + \epsilon_{it,b}. \tag{1}$$

Our dependent variable, $S_{it,b}$ takes a value of one if store i's actual sales on day t fall into bin b, and zero otherwise. Our main regressor of interest, K_0 is a dummy variable that equals one if T is attainable on that day *and* the current sales bin contains the target T. In addition, we include indicators for the five sales bins that are just above or below the target's bin on the same day. Thus, the coefficient β_0 estimates the additional likelihood that a store's sales output falls into bin b when that bin contains an attainable target. The ten coefficients β_{-5} through β_5 identify any excess or missing mass in the output bins surrounding an attainable target. Store fixed effects I_i are included to account for the fact that the number of output bins varies across stores. To measure the effects of the two 'pure' kinks affecting the density of sales, we define regressors analogously for $1.3T$ and $1.6T$, and we re-estimate equation (1) for the analogously defined regressors. The regression sample for a specific threshold type (e.g. $1.3T$) includes its *treatment days* (i.e. the last days on which that threshold is attainable), with essentially all other sales days for that team serving as controls.[31] Estimates of equation (1) for T, $1.3T$ and $1.6T$ are presented in columns 1-3 of Table 3.

5.2 Adding Covariates

One limitation of equation (1) is that it does not account for factors like holidays and the day of the week to affect the sales distribution. While we cannot think of an obvious reason why these factors will be correlated with the presence of an attainable sales target *in a particular bin* on the final day of a month, these factors are highly correlated with total daily sales, so it seems reasonable to control for them. Another limitation of equation (1) is the fact that it does not contain *bin* fixed effects. Thus it implicitly assumes that the baseline sales distribution (i.e. the distribution on days when no threshold is attainable) is uniform.

To address the above issues, we adopt a set of fixed effects that allow each store to have its own baseline sales distribution with unrestricted shape. In addition we allow each store's baseline distribution to have a different shape on each day of the week. Both of these goals are accomplished by giving each store its own set of DOW×bin dummies. Similarly, to account for the substantial improvement in sales during holidays, we control for Holiday×bin fixed effects, where Holiday is defined as a categorical variable that identifies 10 major holidays when people shop heavily in China. This yields the following regression:

$$S_{it,b} = \alpha + \beta_0 \cdot K_0 + \sum_{j=-5}^{-1} \beta_j \cdot K_j + \sum_{j=1}^{5} \beta_j \cdot K_j + \Phi_{iDOW,b} + \Theta_{H,b} + \epsilon_{it,b}, \tag{2}$$

[31] Specifically, the control group includes *all* non-last days plus all last days on which none of the three thresholds are attainable. Appendix C.2 shows that our results are robust to several alternative definitions of control days.

where DOW indexes the days of the week and H indexes ten major holidays.

Like equation (1), equation (2) focuses on the monthly target, T. Separate estimating equations for the two kinks, $1.3T$ and $1.6T$, are defined analogously. Importantly, it is essentially never the case that more than one threshold is within reach on the same day. This is because the thresholds are about 0.3*30 = 9 days worth of sales apart, but a threshold must be within three days of sales to be within reach. All told, the control variables $\Phi_{iDOW,b}$ and $\Theta_{H,b}$ comprise 56,252 store×DOW×bin dummies and 3,971 holiday×bin dummies. Standard errors in equation (2) are clustered by store.

Clustering by stores accommodates a potential concern with estimating a complete distribution of sales for every store: Because sales must fall into exactly one of a store's bins on every day, there will be a mechanical correlation of errors across the bins within every store×day cell. Thus it seems essential to cluster at least at the store×day level. Clustering at the store level should handle this issue, plus more general concerns such as autocorrelation of errors within stores over time. With heterogeneity across time and stores being accounted for in this way, equation (2) again tests whether sales are more likely to fall into a bin that exactly attains a threshold, compared to what we would expect that store×day to produce otherwise. Estimates of equation (2) for T, $1.3T$ and $1.6T$ are presented in columns 4-6 of Table 3.

5.3 Threats to Identification

A natural question concerning our approach is whether the variation we exploit –the presence and location of an attainable threshold on the last day of a month– is 'as good as random'. In this section we describe two plausible sources of non-randomness and their implications for our tests between the standard and alternative models.

5.3.1 Reverse Causation

One potential threat to identification is reverse causation of the following form: An unobserved, persistent shock to store output (such as a multi-day sales promotion) could both bring a threshold within reach (by raising sales on previous days) *and* raise the store's chances of exceeding that threshold on the current day. Such shocks are a concern for our Table 5 regressions, where mean daily sales is the outcome variable; we address this concern in Table 5 with flexible controls for lagged sales. Unobserved, persistent shocks do not, however, pose an obvious threat to our bunching analyses in Tables 3, 4, and 6 for two reasons. First, correlated shocks should not produce gaps or bunching at commission schedule thresholds: they simply predict that sales should be higher after a high-output day. In contrast, Tables 3, 4, and 6 test for gaps or bunching at attainable thresholds, which can appear at *any* level of output (varying from $100 to three days' worth of sales). Second, the effects of positively correlated sales shocks should apply throughout the entire month, whereas the bunching we detect only occurs on final days of the month.

5.3.2 Advance Planning

A second possible concern is advance planning by teams, which would make the team's inherited level of output on the last day an endogenous choice made by each team. Specifically, suppose that –despite the unpredictable nature of daily sales– some teams successfully adopted a forward-looking strategy of tailoring their sales on the last few days of the month in order to achieve a desired level of monthly sales on the final day. If teams engage in this type of forward-looking behavior, our main test between the classical and alternative models of team behavior remains valid.

To see this, consider first a forward-looking classical team, which would want to end the month 'away' from the pure kinks. Such a team should try to position itself so that, before the last day starts, they are in a good position to miss these kinks on the last day. Thus, advance planning gives them additional tools for avoiding kinks. The same reasoning applies to teams that treat kinks as goals: These teams should work to position themselves so that on the morning of the last day they're in a good position to *just attain* either 1.3T or 1.6T. Advance planning gives them additional tools to achieve their objectives also. Thus, our bunching-based test between the standard model and alternative models of team behavior is not invalidated by advance planning behavior.

6 Results

6.1 Main Results

As already noted, estimates of equations (1) and (2) are presented in Table 3. Notably, across all six columns, we find substantial bunching at the target T, and also at the two pure kinks $1.3T$ and $1.6T$. In the absence of controls, column 1 shows that a \$100 sales bin that contains an attainable target T is 17.1 percentage points more likely to be exactly achieved than the same bin when it does not contain an attainable target. In the presence of controls (column 4), this estimated coefficient is 9.2 percentage points. Column 4 also shows statistically significant missing mass in the bin just below the target, equal to 4.7 percentage points. Columns 5 and 6 show estimates from the same regression for coefficients corresponding to the two kinks, $1.3T$ and $1.6T$. These show excess density of 11.4 and 14.9 percentage points respectively. Overall, in all three columns 4 through 6, we find sizable and significant bunching effects at targets and kinks, but –with the exception of the bin just below the target– no strong evidence of nearby missing masses.

To better interpret the excess densities in columns 4-6, it is helpful to have an estimate of the baseline density that we would expect to see in the bins that 'tend to' contain a target or kink, when no target or kink is present. To measure these densities, we use a two-step procedure detailed in Appendix B: First, we calculate for each store the probability the threshold falls into every one of the store's output bins. Then we use these probabilities as weights to calculate the expected density in the output bins where an attainable threshold would 'typically' land on control days. These densities are 10.3, 7.7 and 5.9 percentage points for T, $1.3T$, and $1.6T$ respectively. (Because the higher thresholds are attained less frequently, their baseline densities are lower.) Relative to these baseline densities, the 9.2 percentage points of excess mass at the target T correspond to an 84 percent increase. The excess densities of 11.4 and 14.9 percentage points at $1.3T$ and $1.6T$ correspond to increases of 149 and 252 percent respectively. Thus the estimated effect sizes are substantial in magnitude.

To shed additional light on where the additional density at attainable thresholds in Table 3 *is coming from*, Table 4 estimates the following specification:

$$S_{it,b} = \alpha + \beta_0 \cdot K_0 + \beta_{-5,-1} \cdot K_{-5,-1} + \beta_{1,5} \cdot K_{1,5} + \beta_{\leq -6} \cdot K_{\leq -6} + \beta_{\geq 6} \cdot K_{\geq 6} \\ + \Phi_{iDOW,b} + \Theta_{H,b} + \epsilon_{it,b}, \tag{3}$$

Specifically, we now create a dummy variable, $K_{-5,-1}$ equal to one in all five of the bins just below an attainable threshold, and a dummy variable $K_{\leq -6}$ for all the remaining bins that are further below an attainable threshold. Similarly, we define $K_{1,5}$ and $K_{\geq 6}$ to identify the five bins just above an attainable threshold, and all remaining bins that are beyond the threshold. Notably, a dummy like $K_{\leq -6}$, which represents several underlying output bins, gives the effect of the presence of an attainable threshold on a given day on the probability that output will fall in *one* of those five bins. To estimate the effect of an attainable threshold on the total probability mass in a region represented by such a dummy, the coefficients reported in Table 4 scale up the estimated coefficient estimates by the number of \$100 bins each coefficient represents.[32] When scaled in this way, the coefficients in Table summarize *how the entire sales distribution shifts* on days when an attainable threshold is present, controlling for Bin×Store×DOW and Bin×Holiday fixed effects.

Column 1 of Table 4 shows that the presence of an attainable monthly target, T, on a day creates an extra 9.2 percentage points of density in the exact \$100 bin that contains the target, and an extra 10.2 percentage points of density in the five bins immediately above the target. This extra density above the target is consistent both with the increase in the commission rate at the target, and with teams unintentionally exceeding the target due to the 'lumpiness' of sales. Interestingly, of this 19.4 percentage point increase, 6.9 percentage points come from the five bins just below the target, and 10.3 percent come from even lower output levels.[33] Thus, the target appears to be 'pulling' teams upwards from a substantial distance below it. Estimates for the two pure kinks exhibit similar properties, with much of the extra mass at and above the kinks coming from more than five output bins below. While lumpiness of sales and accelerating marginal incentives likely contribute to these patterns as well, Section 6.3.3 shows that most of the effects in Table 4 are driven by Firm A's largest and highest-selling teams. For these teams, an extra \$500 per day is not an unusual distance to travel.

Finally, to quantify the change in *mean* sales on days when a threshold is attainable, and to verify Table 4's finding of a strong rightward shift in teams' sales distribution when a target or kink is attainable, Table 5 uses a completely different and much simpler estimation approach: We

[32] Please see Appendix B.3 for details.

[33] An insignificant 1.9 percent come from more than \$500 above the target.

simply regress total daily sales on a dummy for whether the team's target (T) or one of its two kinks ($1.3T$ or $1.6T$) were attainable on the last day of a month. Motivated by our concerns about reverse causation from persistent, unmeasured shocks to team sales (see Section 5.3), these regressions include controls for the store's sales in each of the three days preceding the end of the month. While we are less confident in these daily mean sales regressions than in our main estimates, Table 5 provides a useful cross-check of the rightward distribution shift found in Table 4, in part because Table 5's estimation approach does not rely on any 'binning' of store sales.

According to Table 5, sales output is \$240 (41 percent) higher on days where a target is attainable, and \$191 (33 percent) higher when $1.3T$ is attainable, than on comparable days with no attainable threshold. For the next kink at $1.6T$, the point estimate is \$258, but is statistically insignificant. Overall, we view Table 5 as strongly confirming Table 4's finding that the presence of an attainable target *or* kink on a particular day *increases* the level of sales on that day.

6.2 Robustness Checks

6.2.1 Alternative Definitions of Attainability

As part of our heterogeneity analyses, Section 6.3.1 will vary the cutoff of three sales days used to define whether a threshold is *within reach*, and (if today is the last day of the month) whether that threshold is *attainable*. Here, we explore a subtler difference in how attainability is defined. Specifically, in our main analysis, we treated a threshold as attainable if it is within no more than three times the store's mean daily sales during the previous days of the current month. Here, motivated by the large day-of-week effects on sales at Firm A, Appendix C.1 redefines a threshold as attainable if it is within three times the store's mean daily sales during the previous same *days of the week* in the current month. Compared with the original definition, this definition includes some thresholds from high-sales days (like weekends) that were previously considered unattainable, and excludes some observations from weekdays since the attainable range is now defined more narrowly for these days. As shown in Appendix C.1, the bunching effects for T, $1.3T$, and $1.6T$ are robust to this change.

6.2.2 Alternative Control Groups

In our main analysis we maximized the number of control days by including all the non-last days in a month *plus* last days not containing attainable thresholds. In Appendix C.2, we assess four possible concerns with this choice of control groups. One concern is that forward-looking teams behave strategically in the days leading up to the end of the month. While Section 5.3.2 has argued that this is not a source of concern for our claim that teams tailor their effort levels to 'just' attain meaningful monthly sales thresholds, it is nevertheless of interest to assess whether this forward-looking type of behavior might be affecting our results. To address this issue, Tables C.2.1.1 and C.2.1.2 replicate Table 3, first dropping the final week of the month from the control group, then dropping the final three weeks. In both cases, the estimates are very similar to Table 3, suggesting that team behavior in the week leading up to the final day is not qualitatively different from earlier days in the month.

A second possible concern with our baseline control group is that the last days of a month are so unique that no other days from the same month are good comparators. To address this possibility, Table C.2.2.1 replicates Table 3 using *only* last days when no threshold is attainable as controls for last days when such a threshold is attainable.[34] Once again, the results are very similar to Table 3. A third possible concern is that our employees may attempt to bunch at a threshold that is within reach, even on *non-last* days of the month. It is unclear why teams might want to do this, and we see no direct evidence of it. Still, we can eliminate any possible effect of such behavior on our results by using only non-last days in which a threshold is *not* within reach as our controls. We do this in Table C.2.3.1, and find very similar estimates to those in Table 3. Fourth, it seems reasonable to wonder whether the extreme sales variations and target mismatches surrounding the Chinese New Year might have a disproportionate impact on our results (or might even drive all of them). We address this issue in Table C.2.4.1 by excluding observations from January and February from our estimation sample. Reassuringly, the results are again similar to Table 3.

6.2.3 Alternative Bin Widths

In Appendix C.3 we assess the robustness of our results to a different specification choice– bin width. Motivated by simplicity and transparency, our main analysis used a fixed sales bin width of \$100 in all stores for our main regression analyses. To see if this matters, Table C.3.1 replicates columns 4-6 of Table 3 using a bin width that is proportional to each store's expected sales. Specifically, the width of a store's sales bins is set at 0.01 of its monthly target T. Not only does this give us smaller bins in smaller stores, it also adjusts bin width from month to month to reflect the strong seasonality of sales. For example, bins are wider in months containing major holidays, when mean sales are higher. Despite this change, Table C.3.1's results are very similar to Table 3's. In Figure C.3.1, we also replicate a key feature of our initial descriptive analysis – Figure 4— using half-sized bins (0.05T instead of $0.1T$). The pattern is very similar to Figure 4, reassuring us that the raw bunching we observe is robust to other bin widths.

6.2.4 Placebo Thresholds

In Appendix C.4 we probe the statistical significance of our bunching estimates by replicating Figure 5's nonparametric excess density estimates for twelve placebo thresholds, ranging from $0.7T$ through $2.1T$. In contrast to the significant (or almost-significant) bunching at the three actual commission thresholds (p = .000, .000, and .058 for T, $1.3T$, and $1.6T$ respectively), we do not find statistically significant excess density at any of these placebo thresholds. Similar results are

[34] To economize on degrees of freedom, Table C.2.2.1 replaces the Bin×Store×DOW plus Bin×Holiday fixed effects in Table 3 by Bin×Store plus Bin×Holiday fixed effects. (The Table 3 specification has essentially only 12 degrees of freedom and exhibits very high standard errors).

found when we look at excess density in a wider interval that includes one $100 bin on either side of the (actual and) placebo thresholds.

6.2.5 Permutation Tests

In Appendix C.5 we again probe the statistical significance of our results, this time by conducting two types of permutation tests. Our first exercise checks for disproportionate impacts of any individual store on our estimates, by replicating Table 3 while dropping one store at a time. As shown in Figure C.5.1, the main estimated coefficient is similar across all cases for each threshold, indicating that the bunching effects are not driven by a specific store. Second, we address the concern that there might be an insufficient number of treated observations at the $1.6T$ kink, by conducting a randomization permutation exercise recommended by MacKinnon and Webb (2020). To that end, we randomly assign store-month observations to $1.6T$ treatment status. We repeat this procedure 500 times, and plot the distribution of the placebo t-statistics in Figure C.5.2. The 2.5th and 97.5th percentiles are marked in red dashed vertical lines and the correct t-statistic is shown by the blue vertical line. As shown in Figure C.5.2, our estimated coefficients for $1.6T$ are very different from the set of coefficients that are generated by random combinations of store-month observations.

6.2.6 Testing for Advance Planning

While –as we discussed in Section 5.3.2– advance planning of sales effort by teams is not a threat to our main hypothesis test, it is still interesting to know whether such behavior is present. In Appendix C.6, we test for planning behavior by examining the distribution of team sales on *penultimate* days of the month. Specifically, if teams are planning ahead to avoid locating at a kink at the end of the month, we should see missing mass at penultimate-day sales levels that would put a team approximately one day's worth of sales away from the kink. In contrast, we should see excess mass at those points if teams are planning in advance to reach the end of the month 'just' at a kink. Empirically, we do not find any evidence of either excess or missing mass at those points. Appendix C.6 also reminds us that our results are robust to dropping days that are closer to the end of the month (which should the most likely candidates for advance planning activity) from the control group.

6.3 Heterogeneity Analysis

6.3.1 Distance to the Threshold ("Goal Difficulty")

Throughout our analysis so far, we have treated all thresholds that are within three days worth of sales as being within reach. In Table 6, we now disaggregate Table 3's thresholds into three types (using the full specification in columns 4-6):

- 'easy' goals (when the threshold is within 1 day of mean store sales),

- ‘sweet spot’ goals (when the threshold is between 1 and 2 days of sales away), and
- ‘stretch’ goals (when the threshold is between 2 and 3 days of sales away).

If our hypothesis that teams treat commission thresholds as goals is correct, the literature on goals suggests that the thresholds with the largest effects on sales should be the ones that are neither too easy nor too hard to attain (Agarwal et al., 2017). Consistent with these expectations, Table 6 shows that bunching is strongest when a sales target *or* commission kink falls into the ‘sweet spot’. The fact that our teams’ responses to the commission kinks are similar to peoples’ responses to non-pecuniary goals in other contexts confirms our *shared team goals* interpretation of the teams’ behavior.

6.3.2 Team Size

In Section 3.3 we offered two competing hypotheses for how individual workers’ valuation of symbolic rewards could affect the sales distributions of entire teams. In the *uncoordinated maximization* hypothesis, each individual team member directly experiences a symbolic reward when their team passes a threshold in its commission schedule, but the team members all select their effort levels independently. In this model, individual workers’ incentives to work towards ‘just’ attaining thresholds will be rapidly attenuated by free-rider issues and team co-ordination problems as team size grows. Thus we should expect the amount of bunching at both the target and the kinks to diminish with team size. In contrast, the *shared team goals* hypothesis assumes that teams (or their leaders) use the thresholds as convenient, salient goals *in order to* solve co-ordination and moral hazard problems; these goals are communicated to workers and enforced by informal peer monitoring. In this view, bunching can increase with team size, in part because the need for simple but explicit coordination devices is greater in larger teams.

To distinguish between these scenarios, Table D.1.1 replicates Table 4, allowing the effects of thresholds to differ by team size. At both the target and the two kinks we find statistically significant bunching in teams of three and four-plus workers, but no bunching in teams of only two workers. Also, the estimated magnitude of bunching at both the target and kinks increases strongly and monotonically with team size. These findings are confirmed non-parametrically by the raw sales distributions in Figure D.1.1, which replicates Figure 5 by team size. At all three thresholds, Figure D.1.1 shows more bunching in stores of at least 4 workers, compared to smaller stores. Thus, our findings are much more consistent with our *shared team goals* hypothesis (Heath et al., 1999; Allen et al., 2017) than with an *uncoordinated maximization* interpretation of team behavior.

6.3.3 Typical Daily Sales (TDS)

In Table D.1.2, we study effect-size heterogeneity with respect to a store’s typical daily sales volume in the current month. The rationale is that –given the lumpy, stochastic nature of customer traffic– bunching close to desired thresholds should be easier in high-volume store-months. To that end, we first calculate the typical daily sales output of a store-month as $TDS = T/D$, where T is a store's target in the current month, and D is the number of days in the month. Then we present separate estimates for store-months with $TDS <= 500$ and $TDS > 500$. Consistent with our expectations, bunching is more pronounced in store-months with higher levels of expected sales. Also consistent, Table 4's finding of significant shifts in density from more than $500 below the target and kinks is confined to these high-sales team-months. In these situations, improving sales by more than $500 is not as 'big a lift' as in other situations.

6.3.4 Team Experience

If bunching is a behavioral bias that reduces experienced utility (such as a misperceived connection between goal attainment and promotion rates, or incorrect conflation of average and marginal incentives), teams with higher mean tenure at Firm A should have learned *not* to bunch. This is especially the case for the two kinks because standard utility-maximizing behavior involves *avoiding* kinks. In contrast, if –as suggested above– bunching at kinks is an effective way for teams to motivate their members, experienced teams might bunch more than others. As shown in Table D.2.1, the amount of bunching at both the target and the kinks increases with average firm tenure of the sales team; in both cases the interaction terms are large and statistically significant. Once again, the evidence suggests that teams use the kinks as goals to co-ordinate their efforts.[35]

6.3.5 Distance to Head Office

Here we exploit the fact that head office positions –which are never offered to regular sales team members and only rarely offered to sales team leaders– are only offered to team leaders *who live close to Firm A's head office*.[36] Thus, if bunching at kinks is driven by team leaders' (incorrect) beliefs that their team's threshold attainment affects their promotion prospects, bunching at kinks should only occur in stores near the headquarters. To this end, Table D.3.1 replicates Table 3, interacting the target and kink bins with an indicator for whether the store is in the same city as Firm A's headquarters. No significant difference is found, consistent with Firm A's claims that their (rare) promotions of team leaders do not depend on (continuous *or* discretized) measures of the leaders' team sales.

6.3.6 Other Heterogeneity Analyses

In Appendices D.4-D.7, we examine whether our bunching estimates differ along four additional dimensions: recent team turnover events, the type of host institution (department stores or shopping malls), a (pre-analysis) measure of a store's sales volatility, and a store's attainment

[35] Motivated by these findings, we were able to converse with a small number of team members, who confirmed that some team leaders promote the thresholds as informal team goals for the month.

[36] Head office jobs are not attractive to employees living far away, since commuting is infeasible, relocation costs are not covered, and wage increments are relatively small.

of a particular threshold in a recent month. In each of these cases, there are plausible reasons why we might expect each of these dimensions to affect sales bunching. For example, recent turnover might reduce team co-ordination, stores with volatile sales might find it harder to bunch, and stores that have recently achieved a threshold might be more likely to treat it as a reference point. That said, we do not detect any significant mediating effects of these characteristics on our bunching estimates.

6.4 Assessing Alternative Explanations of Our Results

In this section, we discuss four alternative explanations of our unexpected bunching results: unobserved discrete material rewards (UDMRs), ratchet effects, misperception of average versus marginal incentives, and strategic time-shifting of sales. Our analysis marshals three types of evidence: *distributional* results on how the thresholds change the shape of the sales distribution, our *heterogeneity* analyses in Section 6.3, and rich details on the *institutional* environment in Firm A derived from our conversations with management and workers.

6.4.1 Unobserved Discrete Material Rewards

An unobserved discrete material reward (UDMR) is any unobserved aspect of a worker's pay, future prospects, or utility that changes discretely when her team's monthly sales pass one of the three thresholds (T, $1.3T$, or $1.6T$). Importantly, UDMRs do not include connections between these outcomes and *continuous* indicators of a team's sales performance, including practices that link promotions or retention to continuous measures of a team's sales *relative to its targets*. Only unobserved rewards that change discretely when teams pass their commission thresholds can cause the bunching we are trying to explain. We address UDMRs' ability to explain our main results in Appendix E.1; here we provide a brief summary of that analysis.

We start by noting that Firm A views its sales teams primarily as independent, self-managing units, and not as the bottom rung of a corporate hierarchy. More specifically, Firm A's management prefers to minimize its involvement in the daily and monthly operations of the teams, relying instead on strong, team-based commissions to encourage co-operation, problem-solving, and peer oversight *within* each of over 100 teams in widely scattered locations. Consistent with this view, Firm A collects little or no information about the performance of individual workers within the teams: recall that no individual sales data are collected. Also, as further documented below, promotions out of the teams into management are nonexistent for team members, and very rare for team leaders. Overall, career incentives and monitoring of individual sales workers are both minimal, with work incentives relying heavily on the team-based commission.

Within this context, Appendix E.1 considers four channels via which UDMRs could still be tied to team kink attainment. The first of these, *base salaries*, can easily be eliminated because –as noted in Section 2– each worker's base salary is determined by a formula that depends only on their tenure with the firm and their store's location. Second, consistent with a no-layoff policy Firm A has honored since at least 2008, management at Firm A does not *dismiss* individual sales workers for poor performance, or *lay off* individual sales workers for business reasons.[37] Firm A views its no-layoffs / no-dismissal policies as low-cost ways to make commission-based sales jobs more attractive in a tight labor market; for sales workers this implies that neither layoffs nor dismissals by management are a concern.[38] Third, as noted, sales team *members* (who are not leaders) are never promoted to head office jobs. Members are promoted to team *leader* roles only when the current leader departs from the store, typically on the basis of seniority. Leaders receive exactly the same commission pay as other team members, but earn a little more base pay to compensate for duties like scheduling.

Fourth, Appendix E.1.5 considers in detail what we consider to be the most likely way that UDMRs could induce bunching at kinks: the possibility that *team leaders'* promotion prospects (to head office jobs) might be linked to how often their team beats its monthly sales thresholds. Institutional factors that limit such a connection include the fact that these promotions are very rare –only one occurred in our year-long sample period– and are not particularly lucrative.[39] Firm A also claims that these promotions are not based on (either continuous or discretized) team sales; instead, education and people skills are considered more important for these office jobs.[40] Continuing with institutional evidence, even if Firm A's sales team leaders wanted to encourage their members to bunch at kinks to enhance the leader's promotion prospects, leaders have very little material leverage that can be used for this purpose: base salaries are set by a formula, all team members receive the same commission income, and leaders have no discretion over hiring and firing.[41]

Finally, recall that our heterogeneity analyses found that teams located close to Firm A's head office are no more likely to bunch than more-distant teams. However, if team leaders believe that hitting targets raises their promotion chances, we should only see this relationship among team leaders who live close to Firm A's head office. This is because head-office promotions are only offered to team leaders who live nearby.[42] In sum, we do not see a plausible way in which team leaders can –or would wish to– induce their members to 'just' attain commission thresholds in order to enhance their own, minimal promotion prospects.

[37] When Firm A closes a store, all the sales workers are offered jobs at other stores.

[38] Individual sales workers can be encouraged to leave by team members, and we believe this is an operative disciplinary factor in the teams. However, we cannot see any reason why teams would use this pressure to promote bunching at commission kinks, apart from the mechanisms already modeled in the paper (loss aversion, symbolic rewards, and shared team goals).

[39] The vast majority of job openings at head office are filled internally within head office or with applicants from outside the firm.

[40] Consistent with these claims, we note that Firm A does not maintain discretized versions of team sales –such as a team's frequency of threshold attainment– in its personnel and sales accounts: Only the exact sales levels are recorded. While Firm A could, in principle, construct such discretized measures by merging in data on teams' thresholds, its record-keeping practices confirm its claims that such indicators are not used in routine personnel or management decisions.

[41] This lack of leverage is consistent with our estimated effects of team size on bunching: If leaders used leverage over team members to help them bunch at kinks for the manager's benefit, we should see more bunching in smaller teams (where such leverage should be more effective). We find the opposite.

[42] These promotions are not attractive to more-distant workers because relocation costs are not covered and wage increments are relatively small.

Having ruled out any plausible ways that Firm A's *actual* personnel policies link team threshold attainment to measured or unmeasured material rewards, Appendix E.1 next asks whether Firm A's sales workers might mistakenly *believe* that they will receive some discrete material reward if their team exceeds the thresholds in their commission schedule. Here, we point out that a misperceptions-based explanation is inconsistent with the fact that bunching increases with team experience and team size. We also note that there is little reason for Firm A's sales workers to misunderstand or doubt the company's claims about how workers are rewarded. This is because Firm A's personnel policies for sales workers are quite typical of its industry, and because they constitute a *system* of HR policies that fit together well, and are understood as being well suited to this high-turnover labor market.

Summing up, Firm A does not intentionally link any discrete material rewards to team threshold attainment, and is very unlikely to be doing so unintentionally. In addition, several results from our heterogeneity analysis are inconsistent with the hypothesis that Firm A's sales teams falsely *believe* that UDMRs are attached to threshold attainment. Hence, the bunching effect we observe is unlikely to be explained actual or perceived UDMRs.

6.4.2 Ratchet Effects

Ratchet Effects refer to the strategic restriction of output by forward-looking agents, with the goal of preventing the upward adjustment of performance targets in the future by a principal (Charness et al., 2011). We can think of two main challenges that ratchet effects might pose for our analysis. The first concerns the level of our teams' sales targets in 2016, our analysis year. Since the 2016 targets were based on each team's realized sales in 2015, it is reasonable to ask if they were affected by strategic output restrictions in 2015. Such restrictions may, in turn, be correlated with unobserved determinants of team behavior during our sample period (2016). Fortunately, we can easily rule this concern out, because 2016 was the first year (in at least five) in which Firm A tied the sales target to a store's previous year's performance. Importantly, Firm A's decision to link the 2016 targets to 2015 sales was announced after all the 2015 sales were realized. In other words, in our analysis year there was no recent precedent for linking sales targets to the previous year's sales, and the decision to introduce such a link was made too late to allow for any strategic output reduction to affect the 2016 targets.

A more serious consequence of ratchet effects would be if teams believe that their 2017 thresholds will be adjusted upward if their sales during our analysis year (2016) crosses one of the three thresholds. If present, this type of ratcheting could account for bunching at both commission targets and kinks. There are at least three reasons why this is unlikely, however. First and least important is Firm A's high annual quit rate (34%). Under these conditions, the potential future return (in terms of lower sales targets) from sacrificing income today by restricting output will be considerably reduced. The second reason refers to how the teams' 2016 thresholds were calculated: They were a *continuous* function of 2015 output (in fact they were exactly equal to 2015 output).

If the teams believe this practice will continue, there is no reason for the teams to hold their sales below the kinks in our analysis year: Only discrete *jumps* in future thresholds caused by attaining today's threshold can create bunching at today's threshold in a ratchet effects model, and such links have never existed at Firm A.

The final reason relates to our evidence on output distributions. The ratcheting hypothesis, as described above, predicts bunching just *below* a threshold, and a reduction in total density to the right of the threshold. As Tables 3, 4, and 5 demonstrate, this prediction is very different from our findings: increased mass at and above all three thresholds. In other words, the ratchet effect predicts output *restriction* when a threshold is within reach, which is the opposite of what we find.

6.4.3 Misperception of Average versus Marginal Commissions

In a number of contexts, individuals have been shown to confuse average and marginal rewards (De Bartolome, 1995; Ito, 2014). In our context, such individuals would mistakenly believe there was a lump sum bonus attached to the two pure kinks, leading to bunching at all three thresholds. In this section we assess whether this type of misperception –about the structure of the commission schedule itself, as distinct from the misperceived UDMRs discussed in Section 6.4.1– can account for our main results.

Most of the challenges facing the average-marginal misperception hypothesis are the same ones that apply to other forms of misperception, already discussed in Section 6.4.1. These include the fact that in our context the misperceptions would need to be *shared* by an entire work team, and our heterogeneity analyses for team size and team experience. Specifically, contrary to the misperception hypothesis, bunching increases with team size (Section 6.3.2), and with a team's experience at Firm A (Section 6.3.4).

A distinct aspect of average-marginal misperceptions, however, is that –in contrast to contexts like income taxes, where feedback is annual and taxes due when filing have little connection to total annual tax liabilities– consequential feedback on total commission in Firm A is received every month, in each worker's pay statement. These monthly pay slips explicitly calculate the worker's pay from the commission schedule in an easy-to-understand way that clearly demonstrates that there is no discrete increase in total commission income associated with attaining 1.3T and 1.6T. Also, the consequences of misperceptions for the size of the worker's paycheck are direct and immediate.

Finally, while we believe the preceding evidence against average-marginal misperceptions is strong, it is worth asking how we should interpret our results if they were driven by average-marginal misperceptions. Specifically, suppose that adding a pure kink to a commission schedule induces a misperception that average rewards jump upwards there. This 'psychological' reward would have the exact same predictions for the distribution of output as our symbolic reward model. Both of these models represent 'psychological' deviations from the classical one; the main empirical distinction between them is that false perceptions should be mitigated by larger teams and team experience, while effective use of commission thresholds as team goals can plausibly increase with team size and experience.

6.4.4 Strategic Sales Shifting

While we have been interpreting the sales bunching documented in this paper as resulting from workers' effort decisions, bunching could also by caused by strategic changes in the timing of sales, without any overall changes in effort. We open our discussion of this possibility with two observations about sales shifting in our context. The first is that –in contrast to the contexts in which time-shifting has been documented (Larkin, 2014; Benson, 2015)– workers' ability to shift sales timing in our context is extremely limited. Long-term relationships with customers are rare, customers are typically pressed for time and unwilling to postpone sales, pricing discretion is absent, and sales are electronically recorded immediately. Also, customers are particularly averse to short-term over-selling, because products purchased at physical stores are typically non-refundable in China. Second, time-shifting of sales –if it existed– is not necessarily a threat to our claim that teams use the commission thresholds as shared team goals to mitigate free-riding and coordination problems. This is because sales shifting is simply an alternative mechanism a team can use to shift the sales distribution in a given period. 'Classical' teams should shift sales so as to *avoid* kinks, whereas 'non-classical' teams (which treat thresholds as goals) should shift sales to *bunch* at kinks.

With these observations noted, Appendix Table E.2.1 tests for the type of time-shifting that a *non-classical* team would use to just reach a threshold on the last day of a month. Such teams would have an incentive to *'pull' sales into the present* from the future when a threshold is within reach– specifically from the beginning of the next month into the last day of the current month. As shown in Table E.2.1, sales early in a month are *not* lower when the team had an attainable threshold on the last day of the previous month. We conclude that the bunching detected by our regressions in Tables 3, 4 and 6 represents true increases in sales effort on final days of the month, not strategic changes in the timing of sales.

Finally, we consider the possibility of a different form of sales shifting that *could* threaten our claim that teams use commission thresholds as symbolic goals, because this type of shifting would be consistent with the standard labor supply model. In this type of sales shifting, teams near the bottom of a low-marginal-rate segment on the last day of a month would want to *postpone* sales into the following month, in the hope of attaining a higher segment in the following month (Healy, 1985; Larkin, 2014). There are two reasons why this form of time-shifting cannot account for the bunching results in Tables 3, 4 and 6. First, these regressions test for bunching at thresholds that are less than three days of sales away on the last day of a month. Thus, on our treatment days, shifting sales to bunch *just above the bottom of the current commission segment* is not possible: The thresholds we study are typically at least nine days worth of sales apart, so the bottom of the current segment has already been exceeded by a substantial margin.[43] The bunching we test for is

[43]Teams that engage in advance planning could, in principle, arrange to bunch at the bottom of a commission segment at the end of a month by starting to shift sales into the future well in advance of the last day. All our tests for the presence of advance planning are rejected, however (see Appendix C.6).

at the top of the current segment. The second reason relates to our finding that the presence of an attainable threshold on the last day of the month shifts the sales distribution to the right (Tables 4 and 5). This is the opposite of what would happen if the team was shifting sales out of the current day into the following month. We conclude that postponing sales for 'standard' reasons cannot account for the sales bunching we document in this paper.

7 Conclusion

This paper has studied the performance of 103 small retail sales teams that face a nonlinear commission schedule which includes a target (with a bonus attached to it) plus two 'pure' concave kinks. Consistent with many studies of non-linear incentive schemes, and consistent with standard models of worker behavior, we have shown that the presence of an attainable target *or* kink on a workday shifts the distribution of daily sales from below the target or kink towards higher levels. Inconsistent with standard models, however, we have documented a robust and surprising finding: instead of avoiding pure concave (i.e. 'accelerating') kinks in the reward schedule, our teams tend to bunch at those points. Using a combination of institutional evidence, heterogeneity analyses, and distributional evidence (on how the entire sales distribution changes when a kink is attainable), we have argued that the unexpected bunching we observe is best explained by shared symbolic rewards associated with threshold attainment, which are used by teams to motivate and co-ordinate their members. Isolating the symbolic value of thresholds in reward schedules has been challenging because, for bonuses and convex kinks, the standard model and the symbolic rewards models predict the same thing: bunching. Thus, our focus on accelerators –an under-studied feature of compensation policy, where the models' predictions conflict– provides a unique opportunity to isolate the symbolic value of pay schedule thresholds.

From a more pragmatic point of view, our results suggest three broader lessons for incentive design in organizations. The first is to suggest a novel benefit of the piecewise linear reward schedules that are so widely used in workplaces and other contexts. These schedules are often conceptualized as approximations to optimal, continuous schemes, motivated by the need to reduce complexity for agents with limited cognitive capacity (Gjesdal, 1988; Apps et al., 2014). Our results challenge this view by demonstrating that piecewise-linearization itself has additional incentive properties, because it creates thresholds with potential symbolic value. This symbolic value may help explain why piecewise linear schemes are so popular, compared to the continuous schemes they supposedly approximate.

Second, our results illustrate the effectiveness of small, self-managing teams whose pay is tied to team-level performance. While economists once viewed team-based rewards with considerable skepticism, a growing literature (e.g. Knez and Simester (2001); Kuhn and Villeval (2015)) has demonstrated surprising levels of effectiveness. In our case, purely team-based incentives induce over 100 sales teams function to effectively with very little monitoring and oversight from management, filling many managerial functions on their own, including monitoring, discipline and even some aspects of personnel selection (Kuhn and Yu, 2021).

Third, employers who rely on linear commissions and bonuses may wish to consider using accelerators, in addition to or in lieu of those schemes. Compared to those schemes, accelerators have four advantages for employers: First, accelerators are much less costly than raising the value of a linear commission rate because they only apply to sales above the kink, and less costly than bonuses because they only raise *marginal* pay. Second, accelerating commission rates may also have attractive selection properties: they provide a simple, rules-based way to raise the (average *and* marginal) pay of high-ability workers without paying other workers more. Third, at least in our context, most of the negative side-effects of nonlinear incentives, such as timing gaming and price distortion, do not apply. Fourth, as noted our results show that accelerators can create salient thresholds in the pay schedule that can have psychological value to workers, and can be useful as convenient team goals.

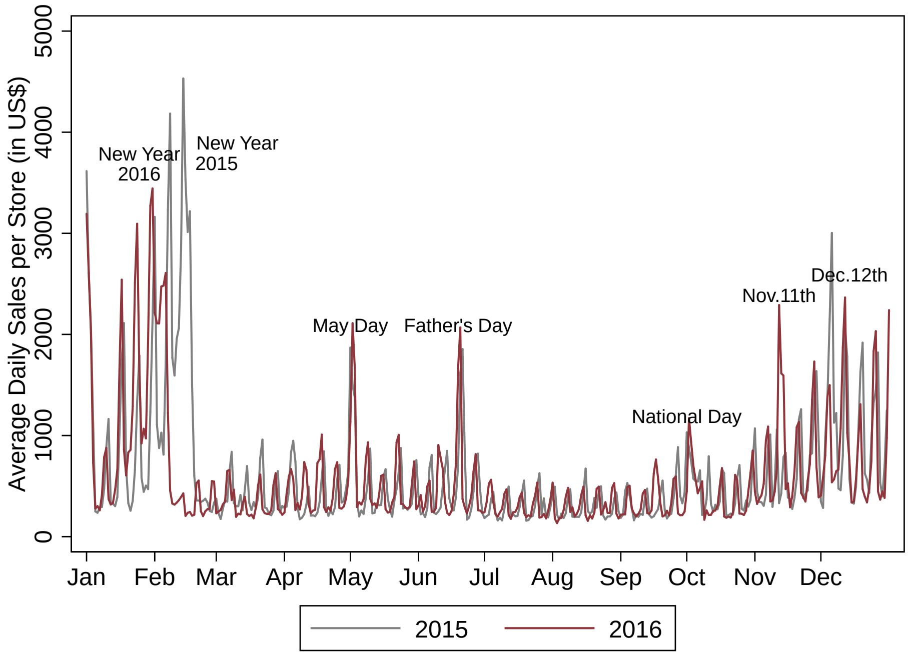
We conclude by reminding readers of two limitations of our analysis. First, if we define Firm A's *accelerator system* as (a) a pair of pure kinks at 1.3T and 1.6T, plus (b) the words attached to the adjacent segments ("Good", "Excellent", and "Superb"), then our main results apply to this *combination* of policies. Specifically, we are confident that Firm A's sales workers derive utility from attaining the kinks in this accelerator system, but we cannot distinguish whether element (a) or (b) is the main source of this utility gain. While distinguishing these aspects is an interesting subject for further research, we hasten to point out that adding attractive words to *any* piecewise-linear reward schedule costs employers nothing.[44] If the words enhance the symbolic value of the thresholds themselves, it is easy to see why these practices are often used together. Second, based on our heterogeneity analyses for team size and experience, our preferred interpretation of the utility value of attaining the thresholds is that teams use these salient output levels as convenient, shared goals to address free-riding and effort co-ordination problems. A stronger test of this mechanism would compare teams' responses to thresholds to those of individual workers. Unfortunately, we cannot conduct this test because Firm A's smallest teams had two workers.

[44] Agarwal et al. (2017) demonstrate psychological value of labelled segments in a setting with no financial incentives, though intrinsic motivation (for resource conservation) may play a greater role in their context than ours.

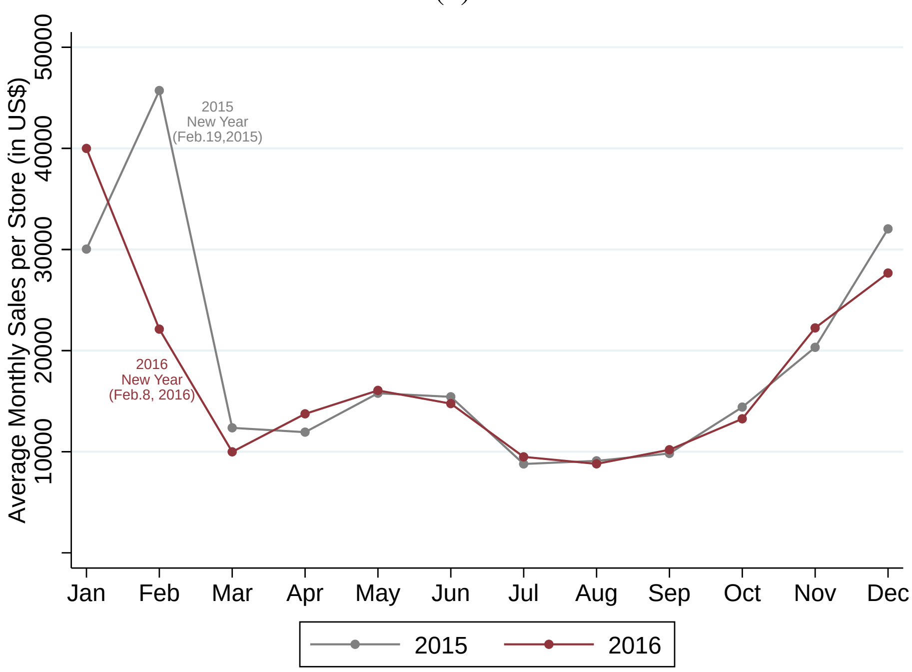
Figures and Tables

Figure 1: Average Daily (Monthly) Sales per Store by Calendar Day (Month)

(a)

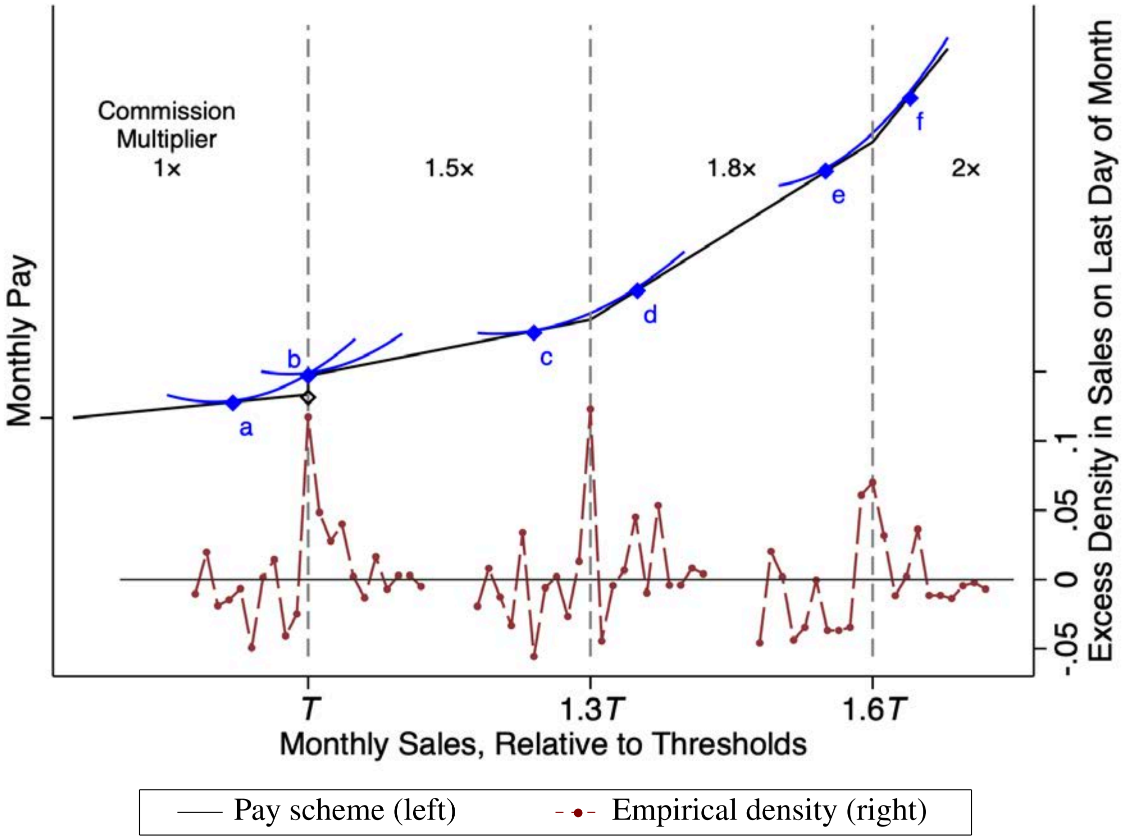


(b)



Notes: Figure 1(a) plots average daily sales per store on every calendar day in 2015 and 2016. The labelled spikes correspond to holidays or major shopping events. New Year denotes the Chinese Lunar New Year. November 11th and December 12th are the major shopping events in China, similar to Black Friday and Cyber Monday in the United States. Figure 1(b) plots average monthly sales per store by calendar month in 2015 and 2016.

Figure 2: Pay Scheme, Standard Model Predictions, and Empirical Excess Densities



Notes: This figure plots Firm A's pay scheme in 2016 (left axis, black line), key indifference curves associated with our theoretical predictions (blue curves), and the empirically observed excess density near the kinks (right axis, dashed brown line). The horizontal axis is the store's monthly output in 2016 relative to the monthly target T, i.e. the same store's monthly output during the same calendar month in 2015.

The commission part of the pay scheme is based on a store's total monthly output, multiplied by a pre-determined baseline commission rate ($1\times$). If the store has met its monthly target T, then each team member is rewarded with a lump sum bonus of $15.63 (100 CNY), and is exempt from a $3.13 (20 CNY) penalty, resulting a jump equivalent to $18.76 (120 CNY) in the compensation schedule. On sales above T, the commission rate is 50% higher than the baseline rate ($1.5\times$). There are another two *pure* kinks in the pay schedule: on sales between $1.3T$ and $1.6T$, the commission rate is 80% higher than the baseline rate ($1.8\times$); on sales exceeding $1.6T$, the commission rate is 100% higher than the baseline rate ($2\times$).

With this pay scheme, theoretical indifference curves for the standard model predict: bunching at the output distribution at point b, a gap in the output distribution between points c and d, and a gap in the output distribution between points e and f.

The three histograms are derived from the data in Figure 5. In each panel of Figure 5 they show the gap between the sales density on last days when a given threshold is within reach, and non-last days when the same threshold is within reach. p-values for equal excess density in the bins containing T, $1.3T$, and $1.6T$ on the last day of the month (compared to non-last days) are .000, .000, and .058 respectively.

Figure 3: A Template of Pay Slip

工资条															
									年　月　日			编号			
序号	姓名	应发									应扣				实发数
		基本工资	绩效工资				绩效奖金	电话补贴	加班工资	小计	绩效扣款	缺勤天数	代扣	小计	
			标准	优良	优秀	杰出									
xx	XXX	2000	1000	450	18		100			3568			-200	-200	3368

月度绩效：　优秀

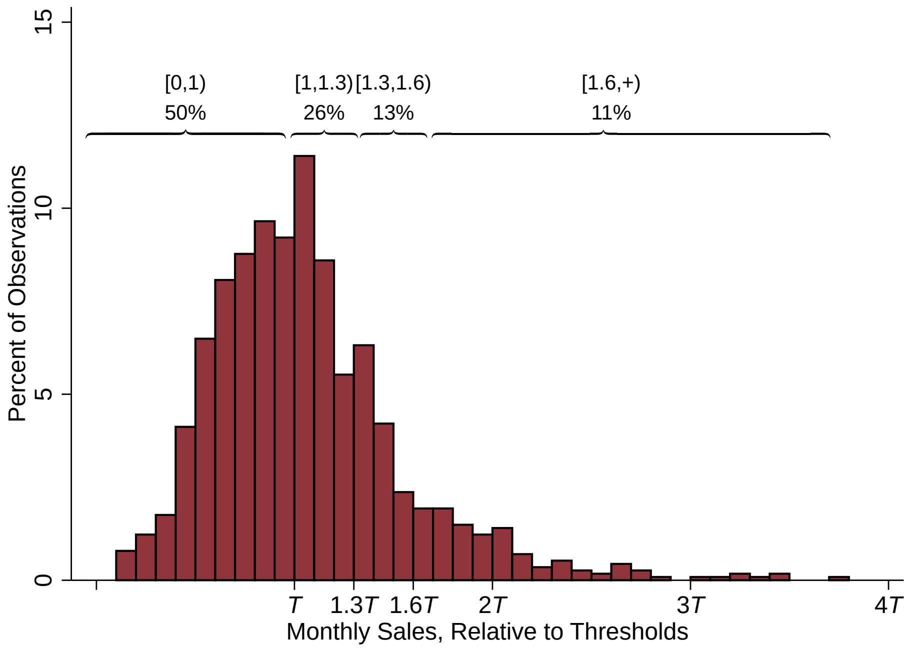
Pay Slip															
									Date				No.		
#	Name	Earnings									Deductions				Net Payment
		Base	Commission				Bonus	Cell Phone Allowance	Overtime Pay	Total	Penalty	Absenteeism	Other Deductions	Total	
			Standard	Good	Excellent	Superb									
xx	XXX	2000	1000	450	18		100			3568			-200	-200	3368

Monthly Asseessment:　Excellent

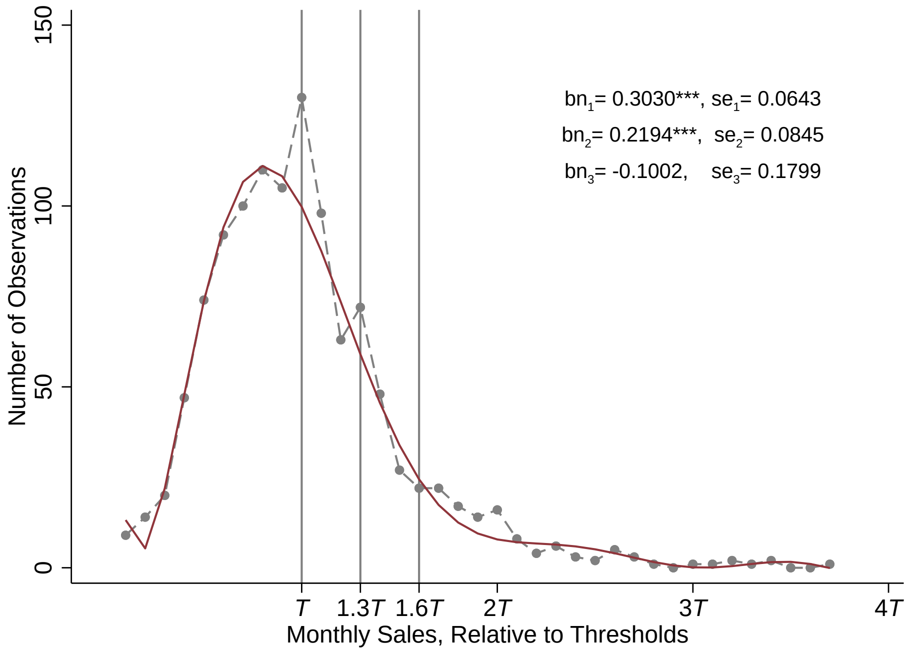
Notes: This figure presents a template of the pay slip to salespeople. The commissions are presented separately with the applicable commission rates, and the ranges are titled as “Standard”, “Good”, “Excellent”, and “Superb”. On the bottom of the pay slip, the monthly assessment takes the highest nonempty title.

Figure 4: Distribution of Monthly Sales, Relative to the Thresholds

(a)

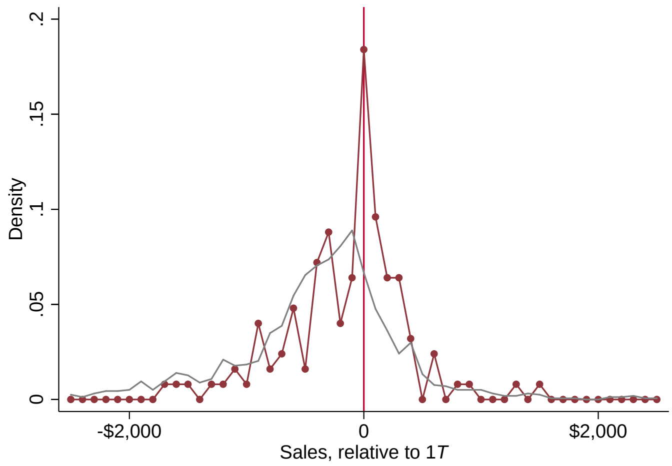


(b)



Notes: Figure 4(a) plots the histogram of relative monthly performance, i.e. the monthly output in 2016 relative to the same store's output in 2015 during the same calendar month (i.e. T). The numbers at the top of the figure indicate the share of all store-month observations falling in each range: e.g. Sales teams fail to reach their monthly target T in 50% of all store-month observations. Figure 4(b) plots the actual number of observations in each bin (gray dashed line), and the estimated number of observations in each bin (red solid curve) from a high-degree polynomial. bn_1, bn_2, and bn_3 indicate the estimated bunching following the method in Chetty et al. (2011) at T, $1.3T$, and $1.6T$ respectively. A test for bunching at $2T$ fails to reject the null (p=0.857).

Figure 5: Distribution of Daily Sales, Relative to the Thresholds



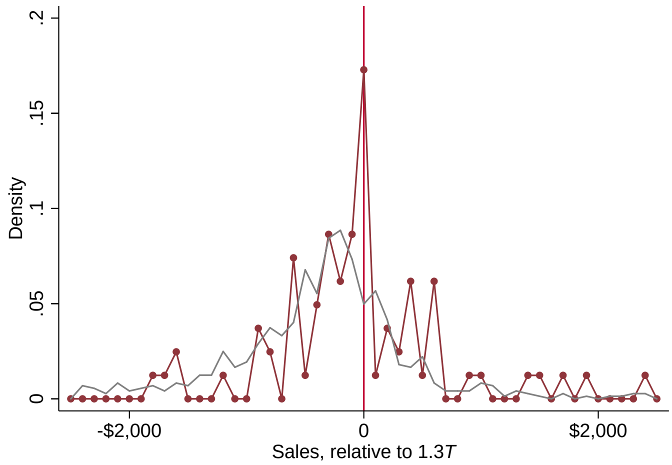
(a) ***T***

H_0: Equal density in Bin 0

t = 4.651; p-value=0.000

H_0: Equal density in Bins -100 $\sim$100

t = 3.533; p-value=0.000



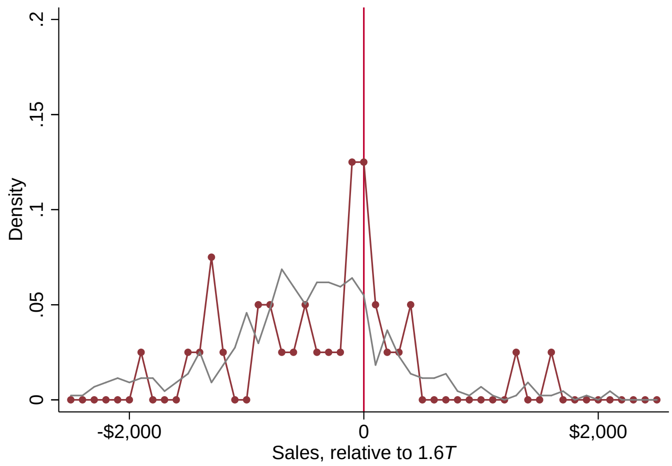
(b)**1.3*T***

H_0: Equal density in Bin 0

t = 4.213; p-value=0.000

H_0: Equal density in Bins -100 $\sim$100

t = 1.780; p-value=0.076



(c) **1.6*T***

H_0: Equal density in Bin 0

t = 1.900; p-value=0.058

H_0: Equal density in Bins -100 $\sim$100

t = 2.980; p-value=0.003

Last days (thresholds within reach) Non-last days (thresholds within reach)

Notes: These figures plot the histogram of excess (or deficient) sales relative to the corresponding thresholds, separately for last days and non-last days. Each figure includes observations when the corresponding threshold is within reach (i.e. a threshold is less than three days' worth of sales away). For better visual display, only the range [-$2,500, $2,500] is presented. Figures (a), (b), and (c) are plotted for kinks T, $1.3T$, and $1.6T$, respectively. The right panel presents t-tests for the equal density hypothesis on last days and non-last days in Bin 0 (or total densities in Bins -100 $\sim$ 100) around each threshold. Figures and tests for placebo kinks are presented in the Appendix C.1.

Table 1: Descriptive Statistics

	Mean	SD	Median	N
Panel A: Product price (in US $)	51.70	49.79	30.94	437
Accessories	16.41	9.77	20.00	20
Shirts and Polos	29.08	10.05	27.81	142
Pants	27.87	10.84	26.25	129
Sweaters	41.63	16.06	40.31	19
Jackets	79.87	34.22	77.81	91
Suits	180.04	53.94	180.94	36
Panel B: Daily sales (in US $)	582	1241	273	34,863
Target size=2	274	368	172	6,376
Target size=3	634	1446	266	18,728
Target size=4	680	1150	382	9,759
Panel C: Relative Monthly Performance	1.07	0.56	1.00	1,143
Target size=2	0.97	0.45	0.91	209
Target size=3	1.10	0.62	1.01	614
Target size=4	1.07	0.49	1.00	320
Panel D: Monthly compensation (in US $)	545	156	512	4,176
Target size=2	507	96	488	446
Target size=3	549	171	513	2,126
Target size≥4	550	145	525	1,604

Notes: Product prices presented in Panel (A) are from a sample of items sold in September, 2016; prices reflect original tag prices and discounted prices.

Target size in Panels (B) through (D) is observed from the annual sales plan, at store-year level. For newly-opened stores whose target size is not available from the annual sales plan, we use their team size 30 days after the opening as the target size for the current year. (The month of store openings are excluded from the analysis, so the first 30 days do not apply.)

In Panel (C), relative monthly performance is a store's monthly output in 2016 relative to the same store's output in 2015 in the same calendar month (i.e. T), for each store-month observations in 2016.

Monthly compensation presented in Panel (D) includes a base salary and a commission component based on team performance, along with the social security payments. Monthly compensations are missing for 6% of employee-month observations.

Table 2: Model Predictions

	Change in Mass		
	Below the Threshold	**At the Threshold**	**Above the Threshold**
A: Predictions at the 'pure' kinks (1.3*T* and 1.6*T*)			
1. Standard Model	–	–	+
2. Symbolic Rewards (SR) Model	–	+	+
3. Loss Aversion (LA) Model (weak loss aversion)	–	–	+
4. Loss Aversion (LA) Model (strong loss aversion)	–	+	+
B: Predictions at the target (*T*)			
1. Standard model	–	+	+
2. Symbolic Rewards (SR) Model	–	+	+
3. Loss Aversion (LA) Model (weak loss aversion)	–	+	+
4. Loss Aversion (LA) Model (strong loss aversion)	–	+	+

Notes:

1. Changes in mass refer to the difference in total density between two situations: (a) when a target or kink is present, versus (b) when the worker faces a linear commission at the (lower) rate that prevails to the left of the threshold.

2. "At the threshold" refers to the bin containing the threshold (and to immediately neighboring bins if teams can't perfectly target). "Below" (above) the threshold refer to the regions of output space outside this interval. Thus, for example, (–,+,+) means bunching at the threshold (and possibly neighboring bins), missing mass below this bunching region, and excess mass above it.

3. The shaded rows indicate that the prediction is consistent with our empirical estimates.

4. Predictions are derived in Appendix A.

Table 3: Estimates of Excess Density in and around Attainable Output Thresholds

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month					
	Specification 1: Raw Estimates			Specification 2: FEs Controlled		
	T	$1.3T$	$1.6T$	T	$1.3T$	$1.6T$
K_{-5}	0.0481	0.0340	-0.0125***	-0.0211	-0.0362	-0.0536***
	(0.0298)	(0.0326)	(0.0012)	(0.0261)	(0.0337)	(0.0125)
K_{-4}	0.0375	0.0463	0.0986	-0.0355	-0.0285	0.0672
	(0.0240)	(0.0328)	(0.0597)	(0.0239)	(0.0250)	(0.0589)
K_{-3}	0.1303***	0.0401	0.0232	0.0396	-0.0331	0.0010
	(0.0396)	(0.0295)	(0.0341)	(0.0380)	(0.0324)	(0.0264)
K_{-2}	0.0568**	0.0935**	0.0178	-0.0290	0.0077	-0.0522
	(0.0285)	(0.0424)	(0.0301)	(0.0262)	(0.0382)	(0.0402)
K_{-1}	0.0321	0.0675**	0.0415	-0.0469**	0.0096	-0.0229
	(0.0329)	(0.0424)	(0.0740)	(0.0184)	(0.0291)	(0.0274)
$\boldsymbol{K_0}$	**0.1707*****	**0.1923*****	**0.1980*****	**0.0919*****	**0.1143****	**0.1494****
	(0.0329)	**(0.0424)**	**(0.0740)**	**(0.0328)**	**(0.0439)**	**(0.0665)**
K_1	0.0944***	0.0357	0.0138	0.0394	-0.0069	-0.0202
	(0.0259)	(0.0237)	(0.0255)	(0.0270)	(0.0228)	(0.0255)
K_2	0.0638**	0.0116	0.0401	0.0274	-0.0239	0.0134
	(0.0262)	(0.0167)	(0.0370)	(0.0250)	(0.0160)	(0.0381)
K_3	0.0333*	0.0116	0.0138	0.0067	-0.0116	0.0008
	(0.0178)	(0.0167)	(0.0254)	(0.0158)	(0.0164)	(0.0235)
K_4	0.0486**	0.0357	0.0401	0.0310	0.0208	0.0352
	(0.0208)	(0.0228)	(0.0349)	(0.0203)	(0.0228)	(0.0349)
K_5	0.0028	0.0116	-0.0125***	-0.0023	0.0051	-0.0187***
	(0.0109)	(0.0165)	(0.0012)	(0.0107)	(0.0172)	(0.0048)
N	2697816	2693899	2691619	2697816	2693899	2691619

$^{*}\ p < 0.1$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

Notes: Columns 1-3 report estimates of equation (1), which controls for store fixed effects. Columns 4-6 report estimates of equation (2), which controls for Bin×Store×DOW and Bin×Holiday fixed effects. The ten Holidays are New Year, Chinese New Year, May Day, National Day, Father's Day, November 11th, December 12th, and three other traditional Chinese holidays. Standard errors reported in parentheses are clustered at the store level. Each column represents a different regression, and the sample in each column comprises the control sample plus the treated sample for the threshold in question (T, $1.3T$, or $1.6T$). An observation is a bin × store × day cell. $\boldsymbol{K_0}$ is an indicator variable, identifying the bin that exactly contains the threshold; the remaining 'K' variables indicate the five bins immediately below and above the threshold.

Table 4: Estimates of Excess Density below, at, and above Attainable Thresholds

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month		
	T	1.3T	1.6T
$K_{\leq -6}$	-0.1033***	-0.1413***	-0.1613***
	(0.0276)	(0.0294)	(0.0602)
$K_{-5,-1}$	-0.0690**	-0.0447	-0.0548
	(0.0344)	(0.0426)	(0.0624)
$\boldsymbol{K_0}$	**0.0919***	**0.1143**	**0.1494**
	(0.0328)	**(0.0439)**	**(0.0665)**
$K_{1,5}$	0.1020***	-0.0170	0.0099
	(0.0385)	(0.0415)	(0.0537)
$K_{\geq 6}$	-0.0192	0.0844**	0.0576
	(0.0192)	(0.0422)	(0.0740)
Bin×Store×DOW	Yes	Yes	Yes
Bin×Holiday	Yes	Yes	Yes
N	2697816	2693899	2691619

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This Table reports estimates of equation (3), with controls for Bin×Store×DOW and Bin×Holiday fixed effects. Standard errors reported in parentheses are clustered at the store level. Each column represents a different regression, and the sample in each column comprises the control sample plus the treated sample for the threshold in question (T, 1.3T, or 1.6T). An observation is a bin × store × day cell. K_0 is an indicator variable, identifying the bin that exactly contains the target or the kink. $K_{-5,-1}$ identifies the five bins *just* below the threshold, and $K_{1,5}$ identifies the five bins *just* above it. $K_{\leq -6}$ identifies all the bins that are below the threshold and more than five bins away, while $K_{\geq 6}$ identifies all the bins more than five bins above the threshold.

Table 5: Daily Mean Sales Regressions

	Dependent variable: Daily Sales (in US$)		
	T	1.3T	1.6T
i. (Last day w/ attainable threshold)	239.7**	191.3**	257.7
	(100.9)	(91.0)	(336.4)
Store×DOW	Yes	Yes	Yes
Holiday	Yes	Yes	Yes
N	34863	34863	34863

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable is the daily sales output; an observation is a store × day cell. The regressor of interest is an indicator variable, taking a value of 1 if the current day is the last day of a month and contains an attainable threshold of the type indicated (T, 1.3T, or 1.6T). All regressions control for the store's sales in each of the three days preceding the end of the month, Store×DOW fixed effects, and Holiday fixed effects. Standard errors reported in parentheses are clustered at the store level.

Table 6: Heterogeneity Examination – Easy Goals, 'Stretch' Goals, and Goals in the 'Sweet Spot'

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month								
	Easy [0, 1)			Sweet [1, 2)			Stretch [2, 3)		
	T	$1.3T$	$1.6T$	T	$1.3T$	$1.6T$	T	$1.3T$	$1.6T$
K_{-5}	-0.0660***	-0.1516**	-0.1366***	-0.0404	-0.0224	-0.0509**	-0.0008	-0.0256	-0.0314**
	(0.0247)	(0.0699)	(0.0277)	(0.0356)	(0.0625)	(0.0222)	(0.0415)	(0.0435)	(0.0122)
K_{-4}	-0.0651**	-0.0759**	-0.0999**	-0.0721**	-0.0249	0.0610	0.0013	-0.0202	0.1189
	(0.0258)	(0.0332)	(0.0445)	(0.0277)	(0.0569)	(0.0906)	(0.0436)	(0.0295)	(0.0962)
K_{-3}	0.0049	-0.0878***	0.1425	0.0216	-0.0876***	-0.0318***	0.0663	0.0234	-0.0161*
	(0.0716)	(0.0271)	(0.1737)	(0.0647)	(0.0200)	(0.0101)	(0.0486)	(0.0650)	(0.0095)
K_{-2}	-0.0819*	-0.1145***	-0.2128***	-0.0082	-0.0055	0.0224	-0.0253	0.0736	-0.0191***
	(0.0467)	(0.0258)	(0.0569)	(0.0552)	(0.0536)	(0.0976)	(0.0222)	(0.0585)	(0.0060)
K_{-1}	-0.0741*	0.0031	0.0148	-0.0678**	0.0200	-0.0603**	-0.0066	0.0059	-0.0258**
	(0.0434)	(0.0682)	(0.0847)	(0.0275)	(0.0532)	(0.0270)	(0.0260)	(0.0336)	(0.0118)
K_0	**0.0735**	**0.0939**	**0.1824**	**0.1188****	**0.1882****	**0.1410***	**0.0848***	**0.0740**	**0.1255**
	(0.0653)	**(0.0853)**	**(0.1127)**	**(0.0597)**	**(0.0843)**	**(0.0829)**	**(0.0447)**	**(0.0582)**	**(0.0920)**
K_1	-0.0067	-0.0391	-0.0099	0.1403**	0.0076	-0.0372***	-0.0137***	0.0144	-0.0164*
	(0.0492)	(0.0465)	(0.0704)	(0.0611)	(0.0414)	(0.0136)	(0.0026)	(0.0345)	(0.0093)
K_2	0.0455	-0.0364	0.0200	0.0189	-0.0331***	0.0504	0.0156	-0.0035**	-0.0217**
	(0.0611)	(0.0442)	(0.0764)	(0.0332)	(0.0066)	(0.0915)	(0.0239)	(0.0017)	(0.0106)
K_3	0.0270	-0.0038	-0.0451***	0.0027	-0.0231***	0.0690	-0.0122***	-0.0100***	-0.0103*
	(0.0369)	(0.0444)	(0.0104)	(0.0235)	(0.0068)	(0.0806)	(0.0028)	(0.0030)	(0.0059)
K_4	0.0231	0.0743	0.1167	0.0550	-0.0141***	-0.0154**	0.0151	-0.0059**	-0.0007
	(0.0370)	(0.0627)	(0.1023)	(0.0393)	(0.0044)	(0.0076)	(0.0237)	(0.0027)	(0.0014)
K_5	-0.0122	-0.0046	-0.0337***	-0.0117***	0.0265	-0.0115**	0.0187	-0.0025	-0.0098
	(0.0228)	(0.0337)	(0.0105)	(0.0025)	(0.0405)	(0.0046)	(0.0232)	(0.0016)	(0.0063)
N	2691267	2689892	2689126	2690917	2689934	2689018	2691086	2689527	2688929

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This Table replicates columns 4-6 of Table 3 for three different types of thresholds– easy, 'sweet spot', and 'stretch'. Easy goals refer to thresholds that are no more than 1 day of mean store sales; 'sweet spot' goals are when the threshold is between 1 and 2 days of mean store sales; and 'stretch' goals are when the threshold is more than 2 days of sales away. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 3 for additional details.

References

Abeler, J., Falk, A., Goette, L., and Huffman, D. (2011). Reference points and effort provision. *American Economic Review*, 101(2):470–92.

Adamson, B., Dixon, M., and Toman, N. (2014). Why individuals no longer rule on sales teams. *Harvard Business Review*.

Agarwal, S., Fang, X., Goette, L., Sing, T. F. S., Schoeb, S. S., Tiefenbeck, V., Staake, T. S., and Wang, D. (2017). The role of goals in motivating behavior: Evidence from a large-scale field experiment on resource conservation. *unpublished manuscript, University of Bonn*.

Ager, P., Bursztyn, L., Leucht, L., and Voth, H.-J. (2022). Killer incentives: Rivalry, performance and risk-taking among german fighter pilots, 1939–45. *The Review of economic studies*, 89(5):2257–2292.

Allen, E. J., Dechow, P. M., Pope, D. G., and Wu, G. (2017). Reference-dependent preferences: Evidence from marathon runners. *Management Science*, 63(6):1657–1672.

Apps, P., Van Long, N., and Rees, R. (2014). Optimal piecewise linear income taxation. *Journal of Public Economic Theory*, 16(4):523–545.

Ashraf, N., Bandiera, O., and Lee, S. S. (2014). Awards unbundled: Evidence from a natural field experiment. *Journal of Economic Behavior and Organization*, 100:44–63.

Benson, A. (2015). Do agents game their agents' behavior? Evidence from sales managers. *Journal of Labor Economics*, 33(4):863–890.

Berg, P., Appelbaum, E., Bailey, T., and Kalleberg, A. L. (1996). The performance effects of modular production in the apparel industry. *Industrial Relations: A Journal of Economy and Society*, 35(3):356–373.

Boning, B., Ichniowski, C., and Shaw, K. (2007). Opportunity counts: Teams and the effectiveness of production incentives. *Journal of Labor Economics*, 25(4):613–650.

Cengiz, D., Dube, A., Lindner, A., and Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. *Quarterly Journal of Economics*, 134(3):1405–1454.

Charness, G., Kuhn, P., and Villeval, M. C. (2011). Competition and the ratchet effect. *Journal of Labor Economics*, 29(3):513–547.

Chetty, R., Friedman, J. N., Olsen, T., and Pistaferri, L. (2011). Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from Danish tax records. *Quarterly Journal of Economics*, 126(2):749–804.

Chevalier, J. and Ellison, G. (1997). Risk taking by mutual funds as a response to incentives. *Journal of political economy*, 105(6):1167–1200.

Cohen, A. (2008). Package size and price discrimination in the paper towel market. *International Journal of Industrial Organization*, 26(2):502–516.

De Bartolome, C. A. (1995). Which tax rate do people use: Average or marginal? *Journal of public Economics*, 56(1):79–96.

de Figueiredo Jr, R. J., Rawley, E., and Shelef, O. (2023). Bad bets: Nonlinear incentives, risk, and performance. *Strategic Management Journal*, 44(1):288–310.

Dobronyi, C. R., Oreopoulos, P., and Petronijevic, U. (2019). Goal setting, academic reminders, and college success: A large-scale field experiment. *Journal of Research on Educational Effectiveness*, 12(1):38–66.

Douthit (1976). The use of sales quotas by industrial firms. *Journal of the Academy of Marketing Science*, 4:467–472.

Freeman, R. B., Huang, W., and Li, T. (2019). Non-linear incentives and worker productivity and earnings: Evidence from a quasi-experiment. *National Bureau of Economic Research, working paper no 25507*.

Friebel, G., Heinz, M., Krüger, M., and Zubanov, N. (2017). Team incentives and performance: Evidence from a retail chain. *American Economic Review*, 107(8):2168–2203.

Genicot, G. and Ray, D. (2020). Aspirations and economic behavior. *Annual Review of Economics*, 12:715–746.

Gjesdal, F. (1988). Piecewise linear incentive schemes. *The Scandinavian Journal of Economics*, pages 305–328.

Hamilton, B. H., Nickerson, J. A., and Owan, H. (2003). Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy*, 111(3):465–497.

Healy, P. M. (1985). The effect of bonus schemes on accounting decisions. *Journal of accounting and economics*, 7(1-3):85–107.

Heath, C., Larrick, R. P., and Wu, G. (1999). Goals as reference points. *Cognitive Psychology*, 38(1):79–109.

Holmstrom, B. and Milgrom, P. (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica*, pages 303–328.

Ito, K. (2014). Do consumers respond to marginal or average price? evidence from nonlinear electricity pricing. *American Economic Review*, 104(2):537–563.

Joseph, K. and Kalwani, M. (1998). The role of bonus pay in salesforce compensation plans. *Industrial Marketing Management*, 27:147–159.

Kaur, S., Kremer, M., and Mullainathan, S. (2015). Self-control at work. *Journal of Political Economy*, 123(6):1227–1277.

Kleven, H. J. (2016). Bunching. *Annu. Rev. Econ*, 8:435–64.

Kleven, H. J. and Waseem, M. (2013). Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan. *Quarterly Journal of Economics*, 128(2):669–723.

Knez, M. and Simester, D. (2001). Firm-wide incentives and mutual monitoring at Continental Airlines. *Journal of Labor Economics*, 19(4):743–772.

Koch, A. K. and Nafziger, J. (2020). Motivational goal bracketing: An experiment. *Journal of Economic Theory*, 185:104949.

Kosfeld, M. and Neckermann, S. (2011). Getting more work for nothing? Symbolic awards and worker performance. *American Economic Journal: Microeconomics*, 3:1–16.

Kuhn, P. and Villeval, M. C. (2015). Are women more attracted to co-operation than men? *The Economic Journal*, 125(582):115–140.

Kuhn, P. J. and Yu, L. (2021). How costly is turnover? evidence from retail. *Journal of Labor Economics*.

Lambrecht, A., Seim, K., and Skiera, B. (2007). Does uncertainty matter? Consumer behavior under three part tariffs. *Journal of Marketing Research*, 26(5):698–710.

Larkin, I. (2014). The cost of high-powered incentives: Employee gaming in enterprise software sales. *Journal of Labor Economics*, 32(2):199–227.

Larkin, I., Pierce, L., and Gino, F. (2012). The psychological costs of pay-for-performance: Implications for the strategic compensation of employees. *Strategic Management Journal*, 33(10):1194–1214.

Lawler, E. and Mohrman, S. A. (2003). Pay practices in Fortune 1000 corporations. *WorldatWork Journal*, 12(4):45–54.

Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review*, 90(5):1346–1361.

MacKinnon, J. G. and Webb, M. D. (2020). Randomization inference for difference-in-differences with few treated clusters. *Journal of Econometrics*, 218(2):435–450.

Mortenson, J. A. and Whitten, A. (2019). Bunching to maximize tax credits: Evidence from kinks in the US tax schedule. *American Economic Journal: Economic Policy*, 12(3):402–432.

MyFirstSalesJob (2020). How to Structure Sales Commission Plans.

Ockenfels, A., Sliwka, D., and Werner, P. (2015). Bonus payments and reference point violations. *Management Science*, 61(7):1496–1513.

O'Connell, W. A. (1989). *Salesforce Compensation,Dartnell's 25th Survey*. Dartnell Corporation.

Oettinger, G. S. (2002). The effect of nonlinear incentives on performance: Evidence from Econ 101. *Review of Economics and Statistics*, 84(3):509–517.

Owan, H., Tsuru, T., and Uehara, K. (2015). Incentives and gaming in a nonlinear compensation scheme. In *Evidence-based HRM: A Global Forum for Empirical Scholarship*. Emerald Group Publishing Limited.

Oyer, P. (2000). A theory of sales quotas with limited liability and rent sharing. *Journal of Labor Economics*, 18:405–426.

Parrinello, R. (2017). Using accelerators to drive winning sales teams. *Workspan*, (August):32–39.

Pierce, L., Rees-Jones, A., and Blank, C. (2022). The negative consequences of loss-framed performance incentives. *NBER working paper*, (26619).

Royer, H., Stehr, M., and Sydnor, J. (2015). Incentives, commitments, and habit formation in exercise: Evidence from a field experiment with workers at a Fortune 500 company. *American Economic Journal: Applied Economics*, 7(3):51–84.

Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy*, 2(3):180–212.

Scudder, M. D. (2017). Will the DOL rescind the tip pool rule? *The National Law Review*.

Seibold, A. (2021). Reference points for retirement behavior: Evidence from German pension discontinuities. *American Economic Review*, page forthcoming.

Shue, K. and Townsend, R. R. (2017). How do quasi-random option grants affect ceo risk-taking? *The Journal of Finance*, 72(6):2551–2588.

Appendices

Contents

A Appendix A: Theory **3**

A.1 Assumptions . . . 3

A.2 Predictions for the Kinks ($1.3T$ and $1.6T$) . . . 4

A.2.1 The Standard Model . . . 5

A.2.2 The Symbolic Rewards Model . . . 5

A.2.3 The Loss Aversion Model . . . 6

A.3 Predictions for the Target (T) . . . 9

B Appendix B: Methods **10**

B.1 Statistical Significance of Bunching in the Monthly Sales Distributions . . . 10

B.2 Calculating Baseline Densities in the Output Bins that 'Typically' Contain the Target and Kinks . . . 10

B.3 Rescale Estimates for Indicators that Include Multiple Bins . . . 11

C Appendix C: Robustness **12**

C.1 Alternative Definition of Attainability . . . 12

C.2 Alternative Control Groups . . . 14

C.2.1 Excluding Days Near the End of the Month from the Control Group . . . 14

C.2.2 Using *Only* Last Days as Controls . . . 17

C.2.3 Excluding *Non-Last* Days with Reachable Thresholds from the Control Group 18

C.2.4 Excluding All Observations from January and February . . . 19

C.3 Alternative Bin Definitions . . . 21

C.4 Bunching Relative to Actual and Placebo Thresholds . . . 24

C.5 Permutation Tests . . . 30

C.6 Advance Planning . . . 33

C.6.1 Testing for Advance Planning . . . 33

C.6.2 Dropping Days that are Most Likely to be Affected by Advance Planning . 36

C.6.3 Implications of Advance Planning . . . 36

D Appendix D: Heterogeneity **37**

D.1 Team Size and Expected Sales . . . 37

D.2 Team Experience Effects . . . 42

D.3 Promotion Prospects . . . 43

D.4 Turnover Events . . . 44

D.5 Host Institution Type . . . 45

D.6 Store Sales Volatility . . . 46

D.7 Previous Threshold Attainment 47

E Appendix E– Alternative Explanations **48**

E.1 Unobserved Discrete Material Rewards (UDMRs) 48

E.1.1 Base Salaries 48

E.1.2 Layoffs and Dismissals 48

E.1.3 Promotions of Team Members 49

E.1.4 Promotions of Team Leaders 49

E.1.5 Workers' Perceptions of Firm A's Personnel Policies 51

E.2 Testing for Sales Shifting 52

A Appendix A: Theory

This Appendix derives the predictions of three models for the effects of adding two forms of non-linearity to a previously-linear cash incentive scheme. The first form of non-linearity is a concave *kink* at which only the commission rate increases, like those at 1.3T and 1.6T in Firm A. The second form of non-linearity occurs at Firm A's *target* (T) where an increase in the commission rate is combined with a small bonus. The **standard model** assumes workers derive utility only from money and leisure, and have quasi-linear utility in money. In the **symbolic rewards (SR)** model, workers derive a lump sum psychic reward from attaining the threshold. In the **loss aversion (LA)** model (based on Kőszegi and Rabin (2006)), the threshold creates a downward discontinuity in the marginal utility of money. In all cases our models treat teams as unitary decisionmakers, maximizing the utility of a representative worker. Thus, while we empirically estimate the effects of factors that might exacerbate free-riding and co-ordination failures –like team size and experience– our theoretical analyses abstract from these issues.

We begin by formally stating the assumptions of each model, then we derive the three models' predictions for the *kinks*. We conclude by deriving the predictions for the *target*, T. All the predictions are summarized in Table 2 of the paper.

A.1 Assumptions

All three models are special cases of the following framework. The utility of a team's representive worker is:

$$u(y, q) = v(y) - \frac{c(q)}{\psi}, \tag{A.1}$$

where y is income and q is output. Marginal costs of producing output, $\frac{c'(q)}{\psi}$ are increasing, convex and differentiable everywhere; $\psi > 0$ represents heterogeneity in these costs. While teams at Firm A face two important sources of cost heterogeneity –the mean ability of their members and the current day's level of customer traffic– to simplify the discussion throughout this Appendix we adopt the more familiar framing where higher levels of ψ represent higher-ability 'workers'. Utility derived from income, or *effective income*, is given by $v(y)$. Working in terms of effective income allows us to express symbolic rewards and loss aversion as equivalent to the following changes to the commission rate, relative to a threshold output level, y^t:

- the lump-sum symbolic award is equivalent to a cash bonus at the threshold.
- loss aversion is equivalent to a convex kink in the commission schedule at the threshold.

In more detail, the **standard model** assumes that:

$$v(y) = y \tag{A.2}$$

The **symbolic rewards** model assumes that:

$$v(y) = \begin{cases} y & \text{if } y < y^t \\ y + B & \text{if } y \geq y^t \end{cases} \tag{A.3}$$

where y^t is the worker's income at the threshold and $B > 0$.

Finally, the **loss aversion** model assumes that:

$$v(y) = \begin{cases} y + \eta\lambda \cdot (y - y^t) & \text{if } y < y^t \\ y + \eta \cdot (y - y^t) & \text{if } y \geq y^t \end{cases} \tag{A.4}$$

where $\eta \geq 0$ represents the strength of *comparison* utility (money relative to the threshold level) and $\lambda \geq 1$ represents loss aversion– the additional marginal utility of money when the team is below the threshold.

Notably, all our models ignore the 'lumpy' and uncertain nature of sales at Firm A, which significantly constrain a team's ability to exactly hit the sales bin that contains a salient threshold. To accommodate this uncertainty, Table 2 focuses on our models' predictions for three fairly coarse outcomes: the amount of probability mass in a small interval around the threshold ("bunching"), and the total amount of probability masses below and above that interval.

A.2 Predictions for the Kinks (1.3*T* and 1.6*T*)

For all three models, this Section compares the predicted behavior of workers under two conditions. In the first condition, workers face a linear baseline commission rate:

$$y = y^K + w_1(q - q^K), \forall q \tag{A.5}$$

In the second, we add an accelerator to this schedule, as follows:

$$y = \begin{cases} y^K + w_1(q - q^K) & \text{if } q < q^K \\ y^K + w_2(q - q^K) & \text{if } q \geq q^K \end{cases} \tag{A.6}$$

where $w_1 < w_2$, and y^K and q^K are the income and output levels at the kink in the budget constraint.

For the baseline (unkinked) budget constraint in the standard model, there is a one-to-one mapping between the ability parameter, ψ, and the worker's optimal sales output, q. If the density of ability, $f(\psi)$ is strictly positive and continuous on $(0, \infty)$, this yields a continuous density of observed output levels. Following Saez (2010), we can then derive predictions for the effect of introducing a concave kink at q^k on the distribution of output levels.

A.2.1 The Standard Model

In the standard model, $v(y) = y$. In effective income terms, the budget constraint and the indifference curve of a (who is indifferent between the two segments of the budget constraint) are shown below:

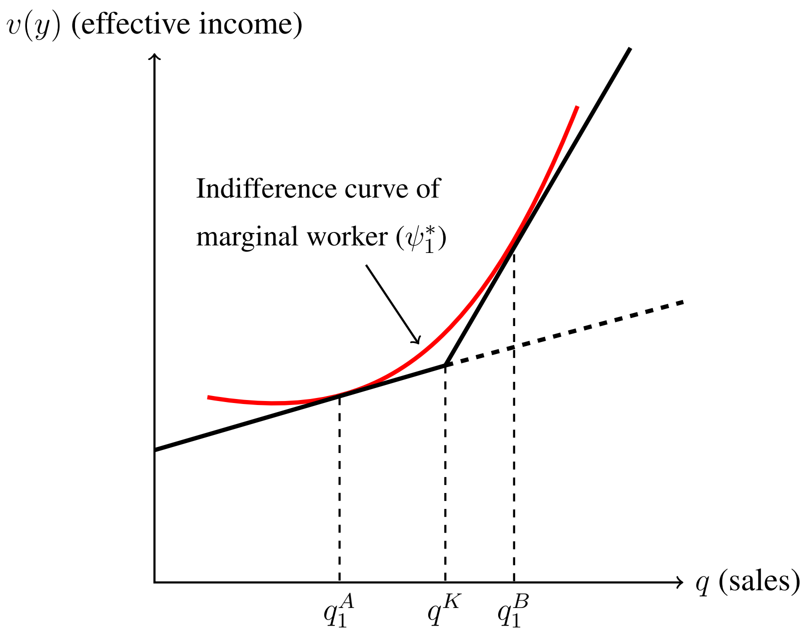


Figure A.1: The Standard Model

In this model, workers with ability (ψ), below ψ_1^* will sell less than q_1^A in the absence of an accelerator, and will not change their output (q) when the accelerator is introduced. Workers with ability above ψ_1^* will raise their (q) to a level above q_1^B. Finally, no workers will locate between q_1^A and q_1^B.

Adding an accelerator therefore:

- leaves the density of sales to the left of q_1^A unchanged;
- reduces the density of sales (to zero) in a region that (strictly) includes the kink (between q_1^A and q_1^B);
- raises the number of workers producing strictly more than the kink (q^K);
- no mass point (bunching) is induced.

A.2.2 The Symbolic Rewards Model

In the symbolic rewards model, $v(y) = y$ if $y < y^K$, and $v(y) = y + B$ if $y \geq y^K$. Now, the effective-income budget constraint and marginal indifference curve take the following form:

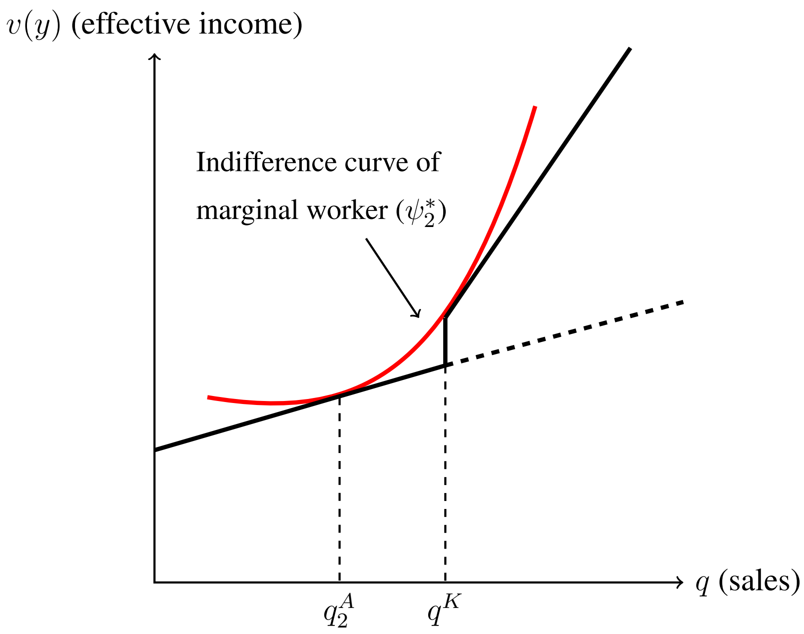


Figure A.2: The Symbolic Rewards Model

Now, workers with ability, ψ, below ψ_2^* will sell less than q_2^A in the absence of an accelerator, and will not change their output (q) when the accelerator is introduced. Because the symbolic reward makes it more attractive to reach the kink, $q_2^A < q^A$. Some of the workers with ability, ψ, above ψ_2* will raise their output (q) to exactly q^K. These are the workers with ability in some interval (ψ_2^*, ψ_2'), where ψ_2' corresponds to tangency (from the right) at q^K.[1]

The remaining workers (with $\psi > \psi_2'$) will raise their output levels to a point strictly above q^K. No workers will locate between q_2^A and q^K.

Adding an accelerator therefore:

- leaves the density of sales to the left of q_2^A unchanged;
- reduces the density of sales (to zero) *below* the kink (between q_2^A and q^K);
- raises the number of workers producing strictly more than the kink (q^K);
- creates a mass point (bunching) at q^K.

A.2.3 The Loss Aversion Model

In this model, $v(y) = y + \eta\lambda(y - y^K)$ if $y < y^K$, and $v(y) = y + \eta(y - y^K)$ if $y \geq y^K$, where $\eta > 0$ measures the importance of comparison utility and $\lambda > 1$ measures the strength of

[1]The indifference curve of the marginal worker will in general be steeper than the budget constraint at q^K; with probability measure zero it will be tangent there, in which case there is no bunching.

loss aversion.[2] Here, the effects of introducing an accelerator depend on the relative size of the loss-aversion parameter (λ) and the jump in the commission rate ($\frac{w_2}{w_1}$) at q^K. There are two cases:

1. *Weak Loss Aversion* ($\lambda < \frac{w_2}{w_1}$):

   When loss aversion is weak relative to the marginal wage increase at the kink, the marginal effective-income gradient to the left of the kink ($\frac{dv}{dq}$) remains lower than the gradient to the right of the kink. Thus, in effective income terms, the worker still faces a concave kink at q^K. Therefore, as in the standard model, the accelerator should create a *gap* in the density of sales around (i.e. on both sides of) the kink point. No bunching of sales should be observed.

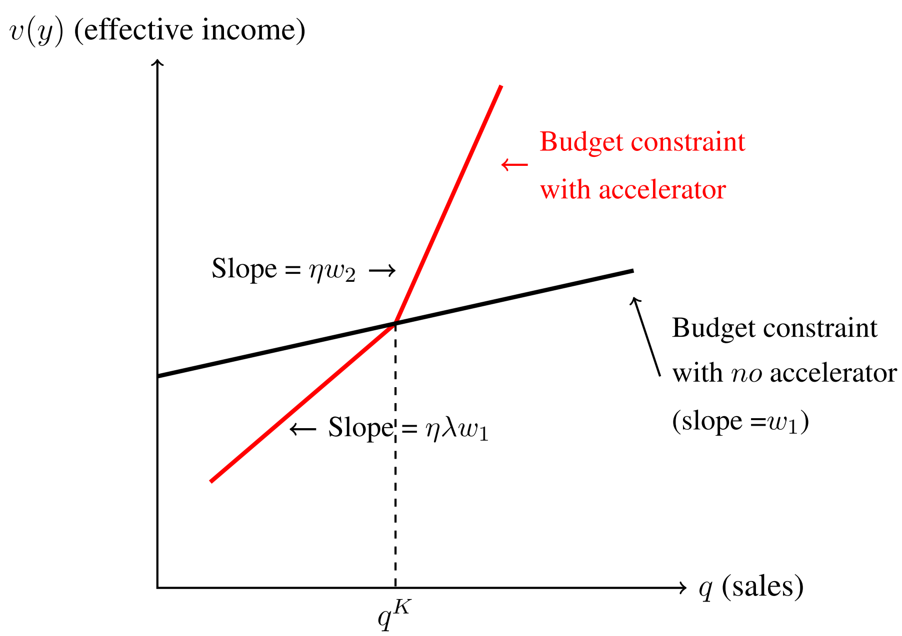


Figure A.3: The Loss-Aversion Model with Weak Loss Aversion

[2]If $\eta = 0$ the worker does not care about income relative to the reference point; if $\lambda \leq 1$ the worker is either loss-neutral or loss-loving.

2. *Strong Loss Aversion* ($\lambda > \frac{w_2}{w_1}$):

   Now the effective-income gradient falls at the kink, creating a convex kink (in effective income terms). By now-familiar reasoning, we therefore expect bunching of sales at q^K:

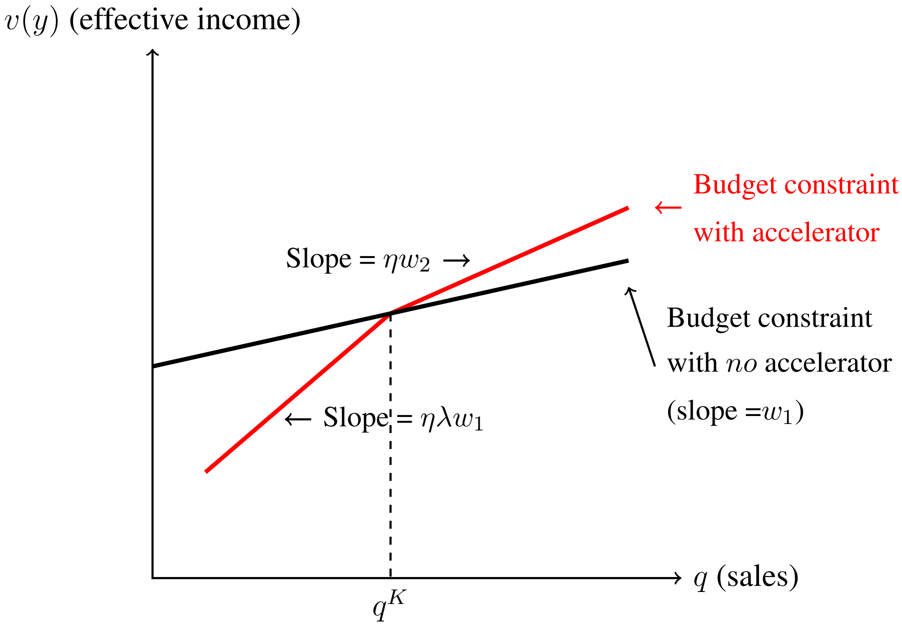


Figure A.3: The Loss-Aversion Model with Strong Loss Aversion

The prediction of bunching at q^K applies to all workers who care about comparison utility ($\eta > 0$), whose loss aversion parameter, λ, exceeds $\frac{w_1}{w_2}$, and who treat some output level, q^K, as a reference point. If we make the additional assumption that workers without a kink in their pay schedule at q^K would not treat it as a reference point (i.e. they would have $\eta = 0$), we can derive the following comparative static predictions about the effect of introducing a commission kink at q^K. Introducing a commission kink:

- creates a mass point (bunching) at q^K;
- reduces the number of teams producing strictly less than q^K;
- increases the number of teams producing strictly more than q^K.

The intuition behind the latter two points stems from the fact that introducing a kink 'activates' the team's comparison utility parameter, shifting it from zero to a positive value. In consequence, the marginal psychic returns to producing additional output increase at all levels of output.[3]

[3]For a similar application of this 'activation' assumption, see Card et al. (2012), who model the effect of providing information about co-workers' wages.

Combining cases 1 and 2, loss aversion can cause bunching at the kink **only if** it is strong enough relative to the increase in marginal cash incentives at the kink. If that is the case, adding an accelerator has the same qualitative predictions for sales density below, at, and above the kinks as the symbolic rewards model.

A.3 Predictions for the Target (*T*)

With the preceding tools in hand, it is straightforward to derive the three models' predictions for the output distribution around the target, T, where an accelerator is combined with a cash bonus. To see this, note first that the standard model is now formally equivalent to the symbolic rewards (SR) model, with the cash bonus now playing the same role as the SR model's *psychic* bonus. Thus, we expect bunching at the target and excess mass to its right. By the same reasoning, the qualitative predictions of the symbolic rewards model are the same at the target as they were at the kinks– the only difference between the target and kinks is that the bonus at the target includes a cash component in addition to the symbolic one. Finally, the presence of a cash bonus means that, in general, the loss aversion (LA) model –like the standard and SR models– *also* predicts some bunching at the target. Extending previous reasoning, this bunching will be accompanied by excess mass to the right of the target when loss aversion is weak, and missing mass if loss aversion is strong.

B Appendix B: Methods

B.1 Statistical Significance of Bunching in the Monthly Sales Distributions

To estimate the significance of bunching in Figure 4(b), we estimate a regression of the following form:

$$c_j = \sum_{i=0}^{p} \beta_i \cdot (z_j)^i + \sum_{i=1,1.3,1.6} \gamma_i \cdot \mathbb{1}[z_j = i] + \epsilon_j, \tag{B.1}$$

where c_j is the number of store-month observations in bin j, z_j is the level of relative performance in bin j ($z_j = 0.1, 0.2, ..., 4$), and p is the order of polynomial. Then we use the predicted values from the above regression to construct the counterfactual bin counts, i.e. $\hat{c}_j = \sum_{i=0}^{p} \hat{\beta}_i \cdot (z_j)^i$. The bunching is then estimated as the excess mass relative to the counterfactual bin counts, i.e. $b_j = c_j - \hat{c}_j = \hat{\gamma}_j$, where j=1, 1.3 and 1.6 in our context. The solid curve in Figure 4(b) shows the counterfactual distribution with an eighth-degree polynomial.

B.2 Calculating Baseline Densities in the Output Bins that 'Typically' Contain the Target and Kinks

Here we describe how we calculate the baseline densities in the output that 'typically' contain the target and two kinks on control days. These are used in Section 5.1 to assess the magnitude of our main estimated coefficients.

We start by computing the mean density in a target's or kink's bin at each store i, separately for T, 1.3T, and 1.6T. For instance, the vector for T at store i, $\boldsymbol{\omega_i^T}$, reflects the likelihood that each bin contains T, i.e. $\boldsymbol{\omega_i^T} = (\omega_{i,1}^T, ..., \omega_{i,N_i}^T)$, where N_i is the number of bins for store i. Then, we restrict to the control sample to compute the mean density of every bin in which store i's sales output falls in, and call it $\boldsymbol{f_i} = (f_{i,1}, ..., f_{i,N_i})$. To compute the likelihood that store i's sales output falls in the bin that contains T, we compute the mean of $f_{i,b}$ across its bins, weighted by $\omega_{i,b}$, i.e. $F_i^T = \boldsymbol{\omega_i^T} \times \boldsymbol{f_i'}$. Thus, F_i^T reflects the likelihood that store i's output falls in the bins where T typically occurs. Finally, we take the sum of F_i^T across all stores to get the baseline density, weighted by the number of observations contributed by each store in the overall sample. This baseline density is estimated to be 10.3 percentage points for the target T. In other words on days when *no* target or kink is attainable, the mix of bins in which the target would typically fall contain 10.3% of the density of store sales. The baseline densities for the two pure kinks 1.3T and 1.6T are defined analogously and equal 7.7 percentage points for 1.3T, and 5.9 percentage points for 1.6T.

B.3 Rescale Estimates for Indicators that Include Multiple Bins

Here we describe how we rescale estimates for indicator coefficients that include multiple bins. These are used in Section 6.1 to identify where the additional density at thresholds is coming from.

The method is illustrated using the threshold T, and we repeat this practice analogously for the thresholds $1.3T$ and $1.6T$. First, we count the total number of bins identified by T, and the number of bins identified by each coefficient. For $K_{<-6}$, this summarizes the number of bins that are more than 5 bins away below the threshold's bin. Using the total number of bins included in T, we calculate the share of bins in the specified range for each coefficient. Then we adjust these shares by comparing them with the share of bins specified by K_0. This gives us the representation scale for each coefficient, adjusted by the representation of the threshold's bin. We then use these scales to multiple the corresponding coefficients and standard errors, to retain the significance levels.

C Appendix C: Robustness

In this Appendix, we assess the robustness of our results to different regression specifications and estimation samples (Appendices C.1 - C.3), and to alternative approaches to statistical inference (Appendices C.4 and C.5). We also assess the robustness of our main claims to the existence of strategic, forward-looking planning by our sales teams (Appendix C.6).

C.1 Alternative Definition of Attainability

In our main analysis, we treated a threshold as attainable if it is within no more than three times the store's mean daily sales during the previous days of the current month. Here, motivated by the large day-of-week effects on sales at Firm A, we redefine a threshold as attainable if it is within three times the store's mean daily sales *during the previous same days of the week* in the current month. Compared with the original definition, this definition includes some thresholds from high-sales days (like weekends) that were previously considered unattainable, and excludes some observations from weekdays since the attainable range is now defined more narrowly for these days. Table C.1.1 replicates Table 3 using this new definition of attainability, with very similar results. For an analysis of robustness to alternative attainability *cutoffs*, please see Table 5 in the paper.

Table C.1.1: Robustness Check – Defining Attainability

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month		
	T	$1.3T$	$1.6T$
K_{-5}	-0.0266	-0.0017	-0.0497***
	(0.0205)	(0.0459)	(0.0120)
K_{-4}	-0.0245	-0.0410**	0.0242
	(0.0233)	(0.0182)	(0.0388)
K_{-3}	0.0293	-0.0247	-0.0022
	(0.0366)	(0.0319)	(0.0226)
K_{-2}	-0.0223	0.0137	-0.0191
	(0.0257)	(0.0341)	(0.0448)
K_{-1}	-0.0306	-0.0028	-0.0231
	(0.0208)	(0.0257)	(0.0245)
$\boldsymbol{K_0}$	**0.0811****	**0.1028****	**0.1312****
	(0.0330)	**(0.0421)**	**(0.0580)**
K_1	0.0331	-0.0194	-0.0192
	(0.0273)	(0.0192)	(0.0226)
K_2	0.0286	-0.0238	0.0094
	(0.0259)	(0.0157)	(0.0347)
K_3	-0.0009	-0.0112	0.0003
	(0.0143)	(0.0161)	(0.0213)
K_4	0.0327	0.0202	0.0310
	(0.0206)	(0.0225)	(0.0319)
K_5	-0.0023	0.0051	-0.0185***
	(0.0111)	(0.0170)	(0.0043)
N	2697816	2693899	2691619

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This Table replicates columns 4-6 of Table 3 using a different definition of attainability. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 3 for additional details.

C.2 Alternative Control Groups

Throughout this paper, we define treated days as last days of a month in which a threshold is within reach: These are the only days on which reaching a specific daily sales level is decisive for attaining one of the three thresholds in the team's monthly commission schedule. In our main analysis, we used the largest possible control group for these treated days: All the non-last days in a month, *plus* the last days on which no threshold was attainable. Since different types of control days have advantages and disadvantages, this Section assesses the robustness of our main results to several different variations in the control group.

C.2.1 Excluding Days Near the End of the Month from the Control Group

We begin with Tables C.2.1.1 and C.2.1.2, which address the concern that forward-looking teams might strategically manipulate their sales *before* the end of the month, in order to facilitate reaching their desired monthly sales on the final day. This strategic behavior could affect sales distributions on control days, with implications for our treatment effect estimates that are hard to assess. To minimize this concern, Tables C.2.1.1 and C.2.1.2 replicate Table 3, excluding days that are closer to the end of the month from our control group: Table C.2.1.1 excludes the last week, and Table C.2.1.2 excludes all three weeks preceding the week's last day. The rationale behind these exclusions is based on the high level of randomness in daily sales in Firm A's stores, which makes advance planning to attain a particular monthly target essentially impossible early in the month, because most of the daily sales shocks have not yet been realized. The estimates in both Tables C.2.1.1 and C.2.1.2 are very similar to Table 3.

Table C.2.1.1: Robustness – Excluding the Week Preceding the Month's Last Day

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month		
	T	$1.3T$	$1.6T$
K_{-5}	-0.0158	-0.0356	-0.0551***
	(0.0283)	(0.0344)	(0.0109)
K_{-4}	-0.0449*	-0.0315	0.0542
	(0.0249)	(0.0260)	(0.0519)
K_{-3}	0.0505	-0.0255	-0.0074
	(0.0399)	(0.0343)	(0.0251)
K_{-2}	-0.0284	0.0054	-0.0580
	(0.0282)	(0.0340)	(0.0387)
K_{-1}	-0.0431**	0.0329	-0.0263
	(0.0191)	(0.0313)	(0.0263)
$\boldsymbol{K_0}$	**0.0940***	**0.0896**	**0.1416**
	(0.0337)	**(0.0427)**	**(0.0622)**
K_1	0.0328	-0.0204	-0.0200
	(0.0264)	(0.0210)	(0.0250)
K_2	0.0260	-0.0245	0.0140
	(0.0262)	(0.0162)	(0.0363)
K_3	0.0003	-0.0079	0.0041
	(0.0146)	(0.0171)	(0.0217)
K_4	0.0326	0.0224	0.0318
	(0.0212)	(0.0230)	(0.0338)
K_5	-0.0024	0.0053	-0.0143***
	(0.0114)	(0.0181)	(0.0045)
N	2150998	2146959	2145790

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Notes: This Table replicates columns 4-6 of Table 3, excluding the week preceding the month's last day from the control group. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 3 for additional details.

Table C.2.1.2: Robustness – Excluding Three Weeks Preceding the Month's Last Day

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month		
	T	$1.3T$	$1.6T$
K_{-5}	-0.0207	-0.0316	-0.0383***
	(0.0278)	(0.0335)	(0.0131)
K_{-4}	-0.0338	-0.0163	0.0529
	(0.0243)	(0.0258)	(0.0611)
K_{-3}	0.0399	-0.0293	-0.0120
	(0.0374)	(0.0332)	(0.0291)
K_{-2}	-0.0244	0.0150	-0.0424
	(0.0281)	(0.0419)	(0.0363)
K_{-1}	-0.0474**	-0.0001	-0.0354
	(0.0200)	(0.0300)	(0.0278)
$\boldsymbol{K_0}$	**0.0978***	**0.1051**	**0.1600**
	(0.0326)	**(0.0440)**	**(0.0695)**
K_1	0.0419	-0.0144	-0.0202
	(0.0275)	(0.0248)	(0.0252)
K_2	0.0240	-0.0259	0.0182
	(0.0258)	(0.0174)	(0.0378)
K_3	0.0038	-0.0124	0.0038
	(0.0159)	(0.0168)	(0.0224)
K_4	0.0257	0.0166	0.0288
	(0.0201)	(0.0224)	(0.0372)
K_5	-0.0081	0.0088	-0.0117**
	(0.0113)	(0.0175)	(0.0048)
N	784450	780533	778253

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This Table replicates columns 4-6 of Table 3, excluding three week preceding the month's last day from the control group. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 3 for additional details.

C.2.2 Using *Only* Last Days as Controls

A different potential concern with our main control group is that last days of a month are so unique that we should not use *any* non-last days as controls. To address this concern, Table C.2.2.1 replicates Table 3, using *only* last days (when *no* threshold is within reach) as controls. Due to the much smaller sample size, we now only control for Bin×Store and Bin×Holiday fixed effects. The estimated effects are also very similar to Table 3.

Table C.2.2.1: Robustness – Using Only Last Days without Attainable Thresholds as Control

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month		
	T	$1.3T$	$1.6T$
K_{-5}	-0.0056	-0.0263	-0.0295***
	(0.0303)	(0.0352)	(0.0108)
K_{-4}	-0.0530*	-0.0044	0.0176
	(0.0272)	(0.0300)	(0.0608)
K_{-3}	0.0721*	-0.0289	-0.0052
	(0.0399)	(0.0361)	(0.0306)
K_{-2}	-0.0047	0.0208	-0.0379
	(0.0271)	(0.0387)	(0.0359)
K_{-1}	-0.0185	0.0297	-0.0334
	(0.0196)	(0.0323)	(0.0291)
$\boldsymbol{K_0}$	**0.1002***	**0.1309***	**0.1343**
	(0.0373)	**(0.0423)**	**(0.0665)**
K_1	0.0271	-0.0306	-0.0026
	(0.0273)	(0.0226)	(0.0231)
K_2	0.0405	-0.0183	0.0233
	(0.0279)	(0.0156)	(0.0339)
K_3	-0.0053	-0.0204	0.0101
	(0.0166)	(0.0201)	(0.0242)
K_4	0.0365	0.0304	0.0294
	(0.0221)	(0.0217)	(0.0330)
K_5	-0.0082	-0.0001	-0.0277***
	(0.0128)	(0.0189)	(0.0085)
N	80506	76467	75298

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This Table replicates columns 4-6 of Table 3, using only last days of the month when there are no thresholds attainable as the control group. Bin×Store and Bin×Holiday fixed effects are controlled due to the smaller sample size. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 3 for additional details.

C.2.3 Excluding *Non-Last* Days with Reachable Thresholds from the Control Group

Another concern is that employees may respond to a threshold that is within reach, even when it is not the last day of a month. While it is hard to understand why teams might do this, Table C.2.3.1 addresses this possibility, by dropping such days from the control group in our Table 3 regressions. The estimates are very similar to Table 3.

Table C.2.3.1: Robustness – Using Only Days without Reachable Thresholds as Control

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month		
	T	$1.3T$	$1.6T$
K_{-5}	-0.0231	-0.0367	-0.0500***
	(0.0261)	(0.0335)	(0.0130)
K_{-4}	-0.0361	-0.0230	0.0696
	(0.0241)	(0.0253)	(0.0600)
K_{-3}	0.0374	-0.0366	0.0023
	(0.0383)	(0.0335)	(0.0264)
K_{-2}	-0.0293	0.0086	-0.0502
	(0.0268)	(0.0381)	(0.0399)
K_{-1}	-0.0460**	0.0091	-0.0232
	(0.0188)	(0.0284)	(0.0276)
$\boldsymbol{K_0}$	**0.0907*****	**0.1145****	**0.1506****
	(0.0325)	**(0.0440)**	**(0.0657)**
K_1	0.0423	-0.0076	-0.0210
	(0.0270)	(0.0231)	(0.0241)
K_2	0.0255	-0.0244	0.0109
	(0.0249)	(0.0159)	(0.0383)
K_3	0.0054	-0.0108	0.0026
	(0.0158)	(0.0160)	(0.0218)
K_4	0.0308	0.0187	0.0314
	(0.0201)	(0.0225)	(0.0358)
K_5	-0.0027	0.0060	-0.0187***
	(0.0109)	(0.0174)	(0.0047)
N	2340212	2336295	2334015

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This Table replicates columns 4-6 of Table 3, using only days when the corresponding threshold is not within reach as control days (i.e. we drop non-last days when the corresponding threshold is within reach from the control group). Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 3 for additional details.

C.2.4 Excluding All Observations from January and February

A final concern with our control groups relates to the fact that the Chinese New Year arrived in February 2015 but in January 2016. As shown in Figure 1(b), this shifted a large volume of sales between these two months. Because a store's 2016 monthly sales targets were equal to its 2015 sales, this led to large, arbitrary shifts in stores' target difficulty between January and February 2016: The January targets were very easy to achieve, and the February targets very challenging.

While these large January-February target shifts are arguably a good natural experiment, it is still useful to know whether they, alone, account for all our results. We address this question in Table C.2.4.1, which excludes all January and February observations from the Table 3 regressions. Once again, the results are similar to Table 3, suggesting that the shift of the Chinese New Year does not account for all our results.

Table C.2.4.1: Robustness – Excluding Observations around the New Year

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month		
	T	$1.3T$	$1.6T$
K_{-5}	-0.0200	-0.0337	-0.0572***
	(0.0288)	(0.0442)	(0.0168)
K_{-4}	-0.0511**	-0.0291	0.1044
	(0.0232)	(0.0315)	(0.0853)
K_{-3}	0.0509	-0.0628*	0.0037
	(0.0417)	(0.0329)	(0.0363)
K_{-2}	-0.0494*	-0.0070	-0.0713
	(0.0250)	(0.0416)	(0.0572)
K_{-1}	-0.0459**	0.0136	-0.0179
	(0.0203)	(0.0341)	(0.0352)
$\boldsymbol{K_0}$	**0.1036*****	**0.1186****	**0.1650****
	(0.0348)	**(0.0506)**	**(0.0794)**
K_1	0.0350	-0.0084	-0.0188
	(0.0287)	(0.0262)	(0.0328)
K_2	0.0313	-0.0264	-0.0115
	(0.0265)	(0.0181)	(0.0360)
K_3	0.0082	-0.0095	0.0081
	(0.0178)	(0.0190)	(0.0310)
K_4	0.0264	0.0209	0.0419
	(0.0213)	(0.0262)	(0.0462)
K_5	-0.0015	0.0079	-0.0240***
	(0.0113)	(0.0194)	(0.0068)
N	2291195	2286735	2284777

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This Table replicates columns 4-6 of Table 3, excluding January and February from the analysis. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 3 for additional details.

C.3 Alternative Bin Definitions

For simplicity and transparency, all our regression analyses used a fixed bin size of $100 across all stores. In Table C.3.1, we assess the sensitivity of our results to this assumption by, instead, using different bin widths for every store. Specifically, Table C.3.1 replicates columns 4-6 of Table 3 after setting the width of every sales bin at 0.01 of each store's monthly target ($0.01T$). The results are very similar to the original Table 3.

In Figure C.3.1 we shift our attention from the regressions to our descriptive analysis, and assess its sensitivity to bin width as well. Specifically, we replicate Figure 4 (the monthly sales distribution) using half-size bins ($0.05T$ instead of $0.1T$). As in Figure 4, the estimated bunching at T and $1.3T$ is significant. At $1.6T$, this figure exhibits positive but insignificant bunching.

Table C.3.1: **Robustness – Defining Distance and Sales Bins by Share of the Target**

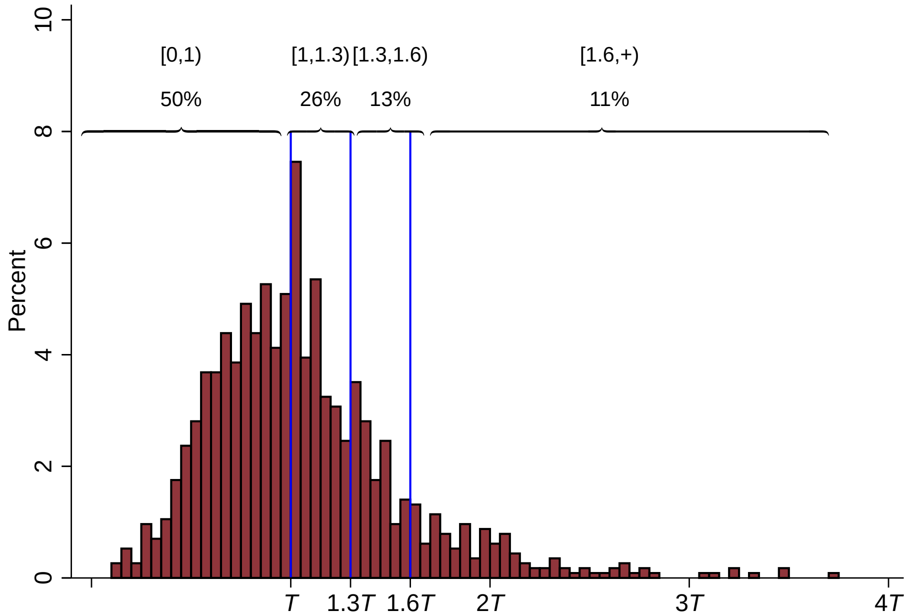
	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month		
	T	$1.3T$	$1.6T$
K_{-5}	-0.0630	-0.0787***	-0.0074
	(0.0407)	(0.0280)	(0.0571)
K_{-4}	-0.0589	-0.0688**	-0.0164
	(0.0379)	(.0306)	(0.0405)
K_{-3}	-0.0620**	-0.0064	-0.0363
	(0.0306)	(0.0402)	(0.0254)
K_{-2}	0.0180	-0.0286	-0.0208
	(0.0384)	(0.0257)	(0.0436)
K_{-1}	-0.0559***	-0.0214	-0.012
	(0.0210)	(0.0329)	(0.0344)
$\boldsymbol{K_0}$	**0.0670***	**0.0791****	**0.1116***
	(0.0342)	**(0.0382)**	**(0.0653)**
K_1	0.0407	0.0049	0.0308
	(0.0302)	(0.0281)	(0.0383)
K_2	0.0321	0.0101	-0.0343***
	(0.0255)	(0.0232)	(0.0068)
K_3	0.0312	-0.0137	0.0252
	(0.0236)	(0.0182)	(0.0338)
K_4	-0.012	0.021	0.0015
	(0.0108)	(0.0243)	(0.0254)
K_5	-0.0031	0.0222	-0.0128***
	(0.0110)	(0.0212)	(0.0035)
N	2292967	2289744	2288239

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

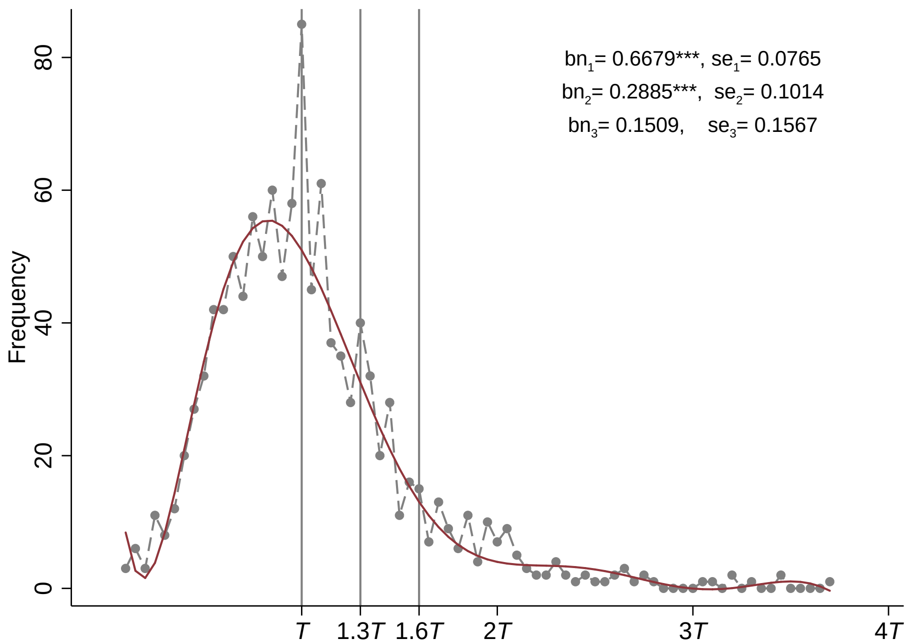
Notes: This Table replicates columns 4-6 of Table 3. Distance and sales are defined into bins according to share of the monthly target T. Each bin is defined as $0.01T$. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 3 for additional details.

Figure C.3.1: Robustness – Distribution of Monthly Sales in Half-Size Bins

(a)



(b)



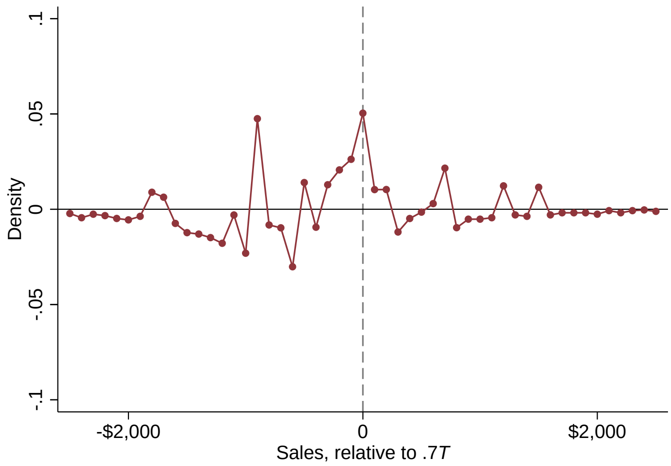
Notes: These figures replicate the practice in Figure 4, where each bin is half the size of bins in Figure 4 (i.e. $0.05T$). Figure (a) plots the histogram of relative monthly performance. Figure (b) plots the actual number of observation (gray dashed line) and the estimated number of observations in each bin (red solid curve) from a high-degree polynomial. The estimated bunching mass at each threshold is reported in the top right corner.

C.4 Bunching Relative to Actual and Placebo Thresholds

In this Appendix, we replicate Figure 5's nonparametric evidence of excess density at attainable commission thresholds (T, $1.3T$, and $1.6T$) in daily sales distributions, for a wide range of placebo thresholds, ranging from $0.7T$ through $2.1T$. To make the patterns easier to see, we show only the *excess* densities at each point (i.e. the density on last days when the threshold is attainable minus the density on non-last days when the same threshold is attainable) rather than those two densities separately. Hypothesis tests for zero excess density in the focal bin, and in the three bins that are centered on the focal bin, are provided. For comparison, we include the actual thresholds (T, $1.3T$, and $1.6T$, shown in red) as well as the placebo thresholds.

Notably, we do not find statistically significant excess density in "Bin 0" (the exact location of the placebo bin) for any of the 12 placebo thresholds, while we find significant (or almost-significant) bunching at the three actual commission thresholds (p = .000, .000, and .058 for T, $1.3T$, and $1.6T$ respectively). While bunching exactly at $1.6T$ is not quite significant at conventional levels, bunching becomes significant if we include the two bins on either side of $1.6T$ (p=.003). At these very high sales thresholds, it may be easier for teams to 'just miss' a threshold they are striving to reach.

Figure C.4.1 Excess Sales Relative to Actual and Placebo Threshold from 0.7T to 2.1T



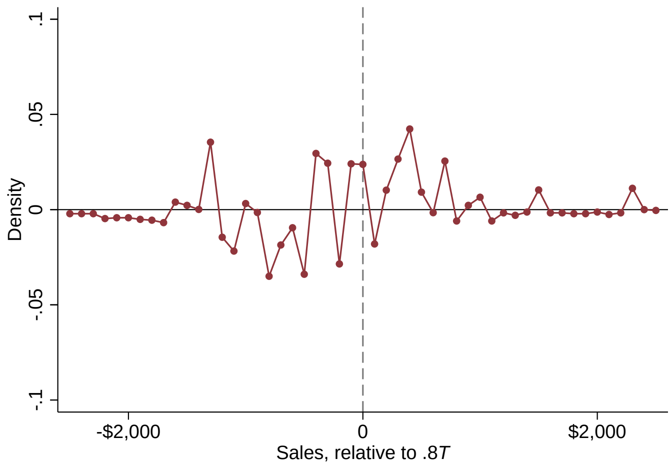
(a) ***0.7T***

H_0: Equal density in Bin 0

t = 1.612; p-value = 0.107

H_0: Equal density in Bins -100 $\sim$ 100

t = 1.699; p-value = 0.089



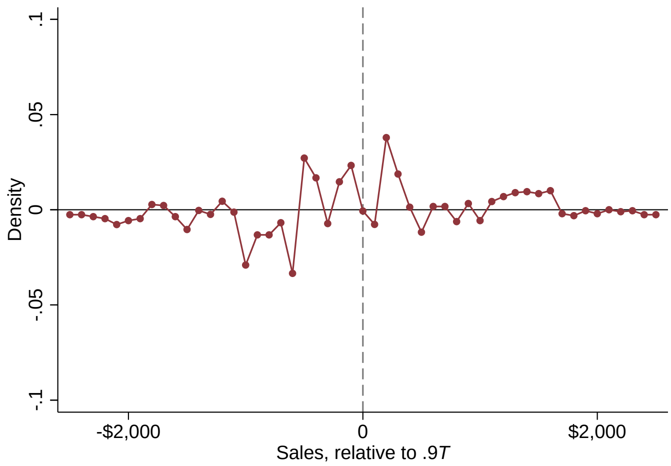
(b) ***0.8T***

H_0: Equal density in Bin 0

t = 0.814; p-value = 0.416

H_0: Equal density in Bins -100 $\sim$ 100

t = 0.476; p-value =0.634



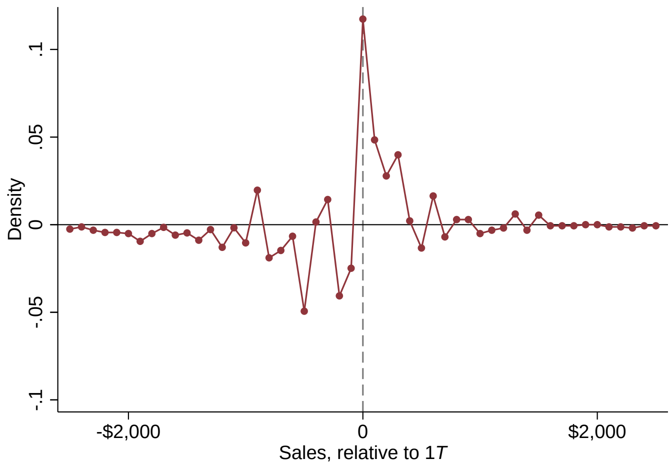
(c) ***0.9T***

H_0: Equal density in Bin 0

t = 0.045; p-value = 0.964

H_0: Equal density in Bins -100 $\sim$ 100

t = 0.322; p-value = 0.748



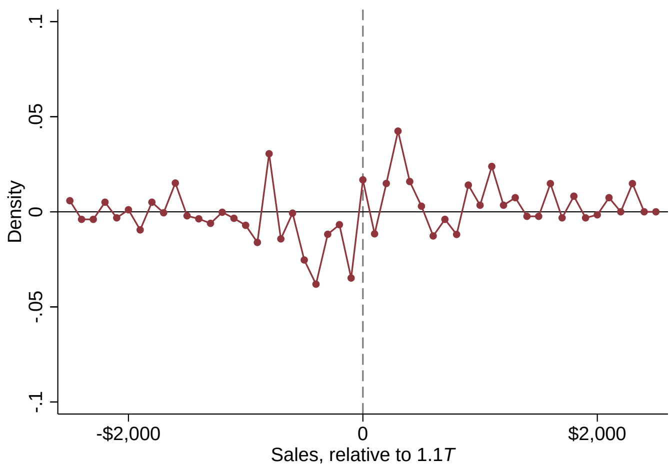
(d) ***T***

H_0: Equal density in Bin 0

t = 4.651 ; p-value = 0.000

H_0: Equal density in Bins -100 ∼ 100

t = 3.533 ; p-value = 0.000



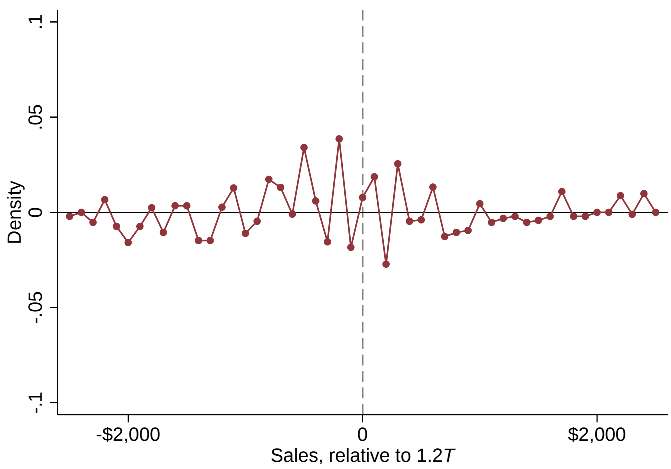
(e) **1.1*T***

H_0: Equal density in Bin 0

t = 0.659; p-value = 0.510

H_0: Equal density in Bins -100 ∼ 100

t = 0.819; p-value = 0.413



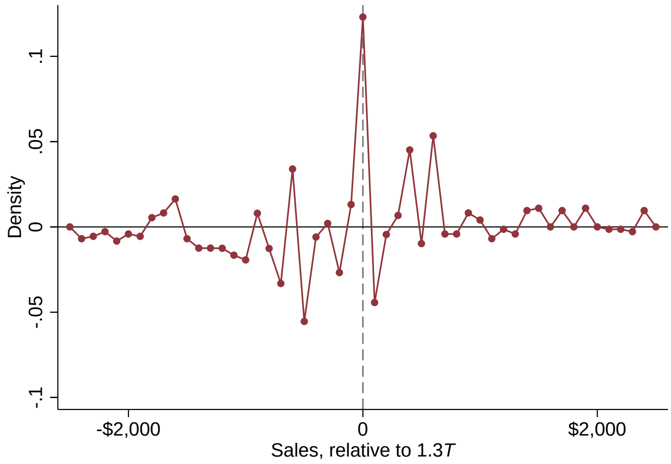
(f) **1.2*T***

H_0: Equal density in Bin 0

t = 0.303; p-value = 0.762

H_0: Equal density in Bins -100 ∼ 100

t = 0.128; p-value = 0.899



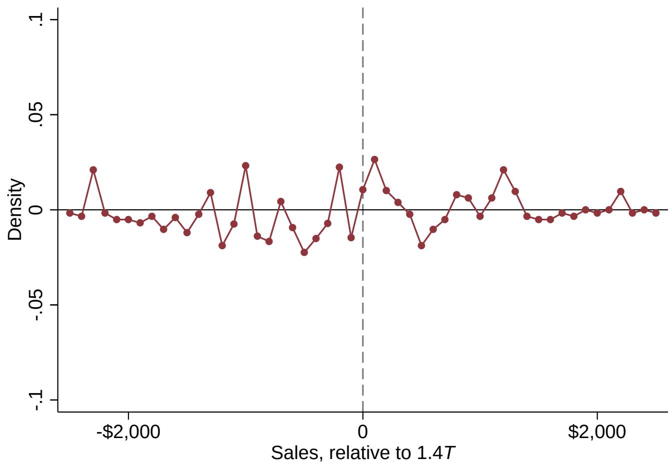
(g) **1.3*T***

H_0: Equal density in Bin 0

t = 4.213; p-value = 0.000

H_0: Equal density in Bins -100 $\sim$ 100

t = 1.780; p-value = 0.076



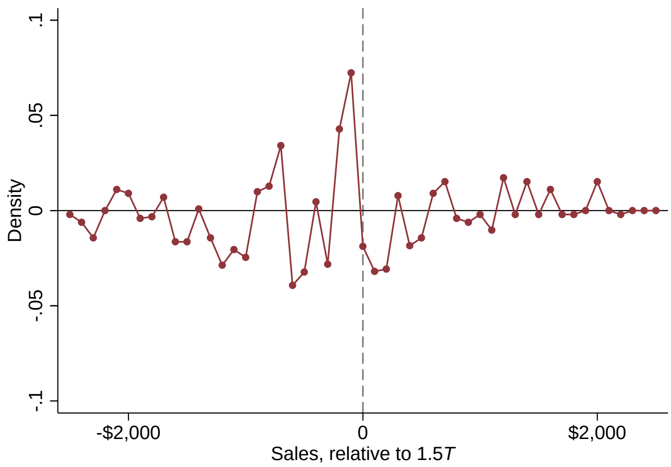
(h) **1.4*T***

H_0: Equal density in Bin 0

t = 0.556; p-value = 0.579

H_0: Equal density in Bins -100 $\sim$ 100

t = 0.756; p-value = 0.450



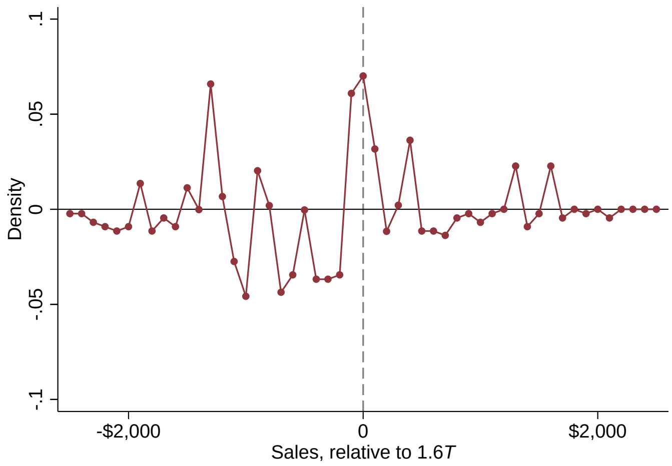
(i) **1.5*T***

H_0: Equal density in Bin 0

t = 0.456; p-value = 0.648

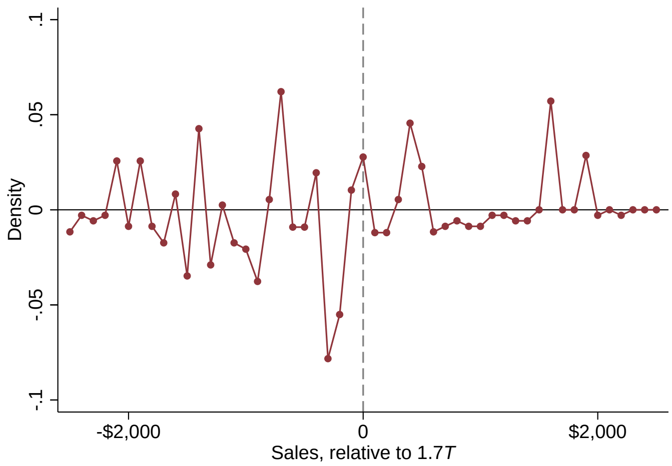
H_0: Equal density in Bins -100 $\sim$ 100

t = 0.793; p-value = 0.428



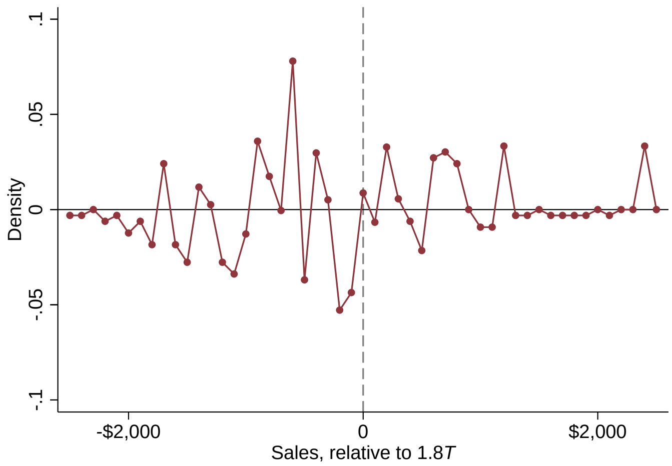
(j) **1.6*T***

H_0: Equal density in Bin 0
t = 1.900; p-value = 0.058
H_0: Equal density in Bins -100 $\sim$ 100
t = 2.980; p-value = 0.003



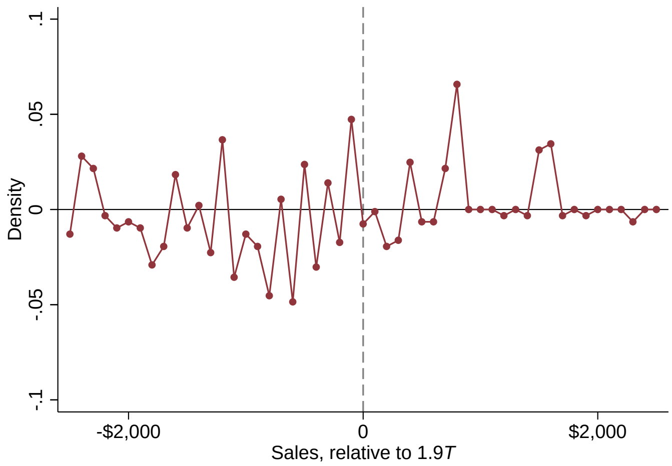
(k) **1.7*T***

H_0: Equal density in Bin 0
t = 0.751; p-value = 0.453
H_0: Equal density in Bins -100 $\sim$ 100
t = 0.542; p-value = 0.588



(l) **1.8*T***

H_0: Equal density in Bin 0
t = 0.271; p-value = 0.787
H_0: Equal density in Bins -100 $\sim$ 100
t = 0.669; p-value = 0.504



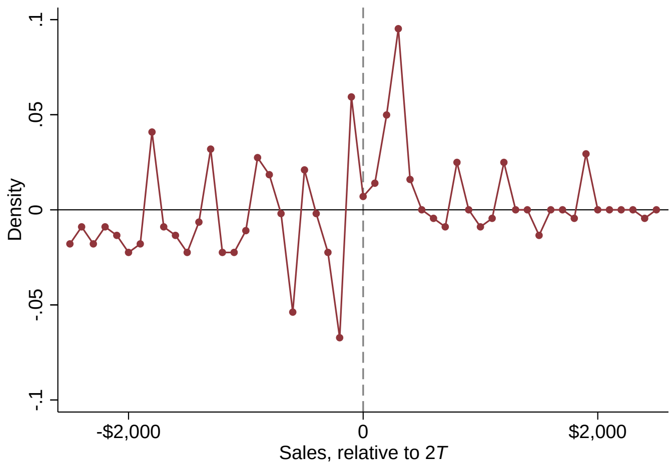
(m) **1.9*T***

H_0: Equal density in Bin 0

t = 0.157; p-value = 0.875

H_0: Equal density in Bins -100 $\sim$ 100

t = 0.634; p-value = 0.527



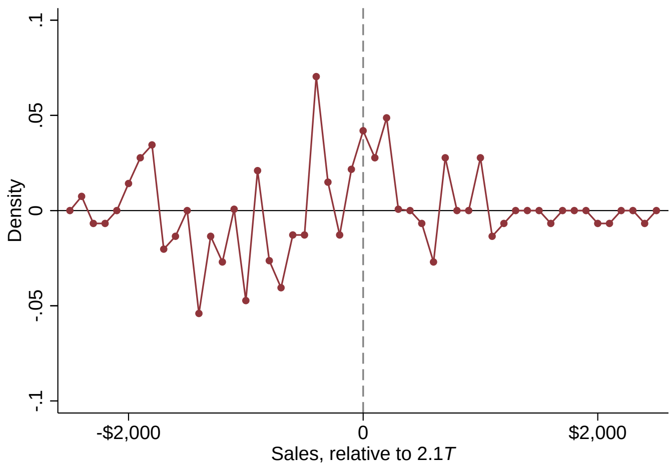
(n) **2*T***

H_0: Equal density in Bin 0

t = 0.201; p-value = 0.841

H_0: Equal density in Bins -100 $\sim$ 100

t = 1.141; p-value = 0.255



(o) **2.1*T***

H_0: Equal density in Bin 0

t = 1.124; p-value = 0.263

H_0: Equal density in Bins -100 $\sim$ 100

t = 1.511; p-value = 0.133

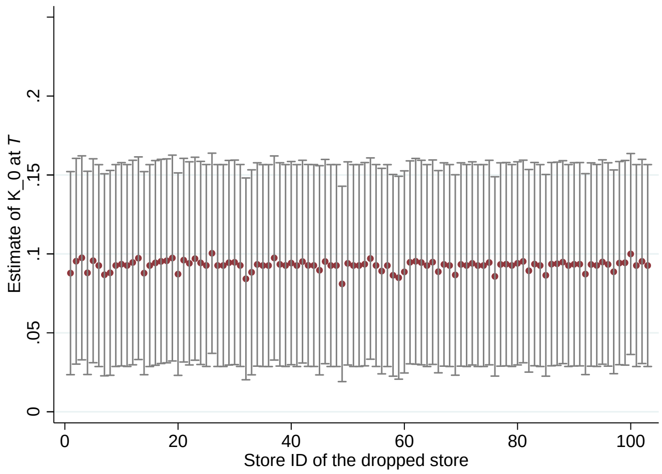
C.5 Permutation Tests

In this section, we perform two types of permutation tests to continue to probe the performance of our statistical inference methods. First, we re-estimate our main regressions dropping one store at a time, to see if any single store accounts for all the clustering. As shown in Figures C.5.1 (a)-(c), the estimated K_0 coefficients are very similar for each corresponding kink, indicating that the bunching effects are not driven by a specific store.

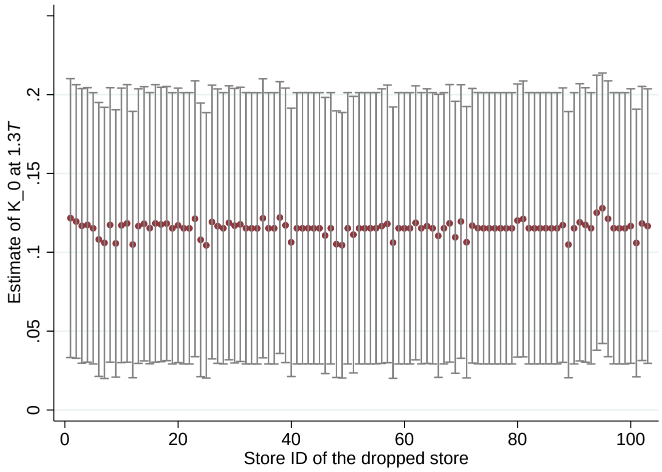
Another concern is that there might be insufficient treated stores at the $1.6T$ kink. To address the "few treated clusters" concern, we conduct a randomization permutation exercise recommended in MacKinnon and Webb (2020). We randomly assign store-month observations to $1.6T$ treatment status and re-estimate Equation (2) to obtain placebo estimates. We repeat this procedure 500 times and report t-statistics of the placebo estimates. In Figure C.5.2, we plot the distribution of the placebo t-statistics, marking the 2.5 and 97.5 percentiles with red dashed vertical lines. The solid blue vertical line is the t-statistics when the 1.6T treatment variables are assigned correctly. As shown in the figure, it is statistically unlikely to observe our main estimated results using other combinations of store-month observations.

Figure C.5.1: Robustness Check – Permutation Test Dropping One Store at a Time

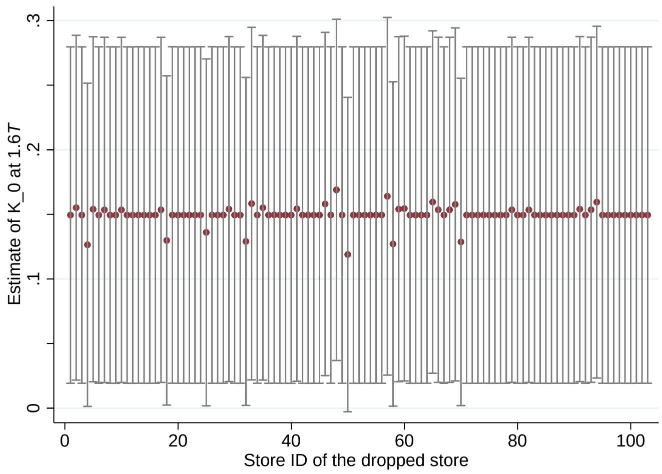
(a)1.0T



(b)1.3T

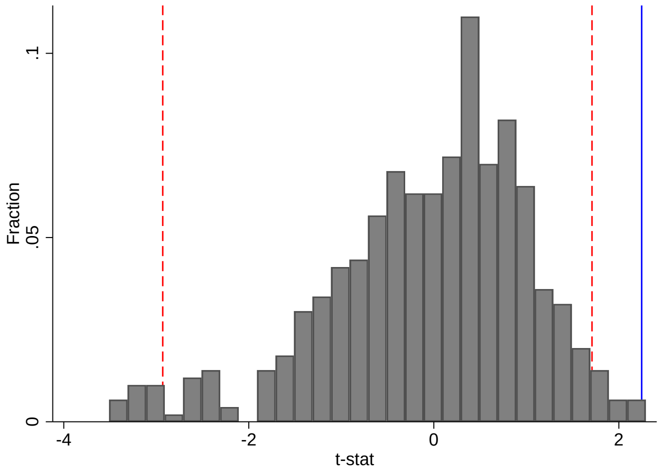


(c)1.6T



Notes: These figures present the estimated coefficients and confidence intervals of a permutation test by dropping one store at a time to re-estimate our main regressions in Table 3.

Figure C.5.2: Robustness Check – Randomization Permutation Test



Notes: This figure plots the distribution of the t-statistics of the placebo estimates, following a randomization permutation practice in MacKinnon and Webb (2020). The red dashed vertical lines represent the 2.5 and 97.5 percentiles of the placebo t-statistics. The solid blue vertical line is the t-statistics when the $1.6T$ treatment variables are assigned correctly.

C.6 Advance Planning

In Firm A, the sales workers are informed of their team's monthly sales goal at the beginning of every month, by their team leader. During the month, the team leaders have continuous access to their team's monthly cumulative sales totals, but the leaders typically only look up these totals at certain key points, such as the middle of the month and the beginning of the final week. Team leaders are free to use this information any way they wish, for example by providing a progress report, or by encouraging members to work towards (or beyond) *any* goal, such as a round number of sales by the fifteenth of the month.

In this environment –at least if daily productivity shocks are small enough– sophisticated teams might find it worthwhile to form a (presumably state-contingent) daily effort plan for the entire month, in order to optimize most effectively against Firm A's nonlinear commission schedule. For example, 'classical' teams –who should want to avoid ending the month at a commission kink– could plan a path of daily effort that finishes the month away from the kinks. Similarly, a more 'behavioral' team that 'wanted' to bunch at a monthly threshold should choose an entire daily effort strategy to 'just' meet a desired sales threshold on the month's last day. In either case, the presence of an attainable kink on a month's last day would be affected by endogenous team decisions that were made earlier in the month.

While we think both these scenarios are unlikely, we do three things in this Section to assess their impact on our main estimates. First, we conduct a test for advance planning based on cumulative sales on the second-last day of the month. If teams plan ahead with a particular goal in mind, this behavior should be easiest to detect on those days. Second, we remind readers of Appendix C.2, where we dropped days that are most likely to be affected by this kind of advance planning from our analysis sample. (These are non-last days near the end of the month.) Our results are unaffected. Third, we discuss how advance planning –if it occurred– would affect our ability to distinguish between the classical model of team behavior and the more 'psychological' alternatives proposed in our paper.

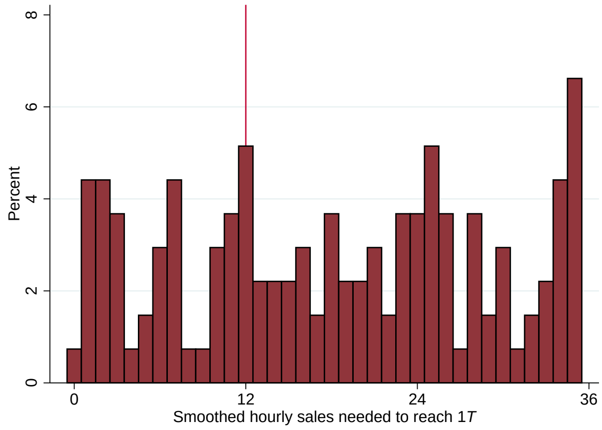
C.6.1 Testing for Advance Planning

In Figure C.6.1, we look for evidence of advance planning by focusing on teams' cumulative monthly sales at the end of the *second*-last day of every month. The idea behind our test is simple: If a psychologically-motivated team plans ahead to 'just reach' a particular threshold on the last day of the month, it should position itself such that it is approximately one normal day of sales away from that threshold on the morning of the last day. In other words, on penultimate days we should see bunching at cumulative sales levels that are one day of typical sales away from a monthly threshold. Conversely, 'classical' forward-looking teams should plan ahead to *avoid* hitting monthly thresholds; in that case we should see missing density at the same sales levels.

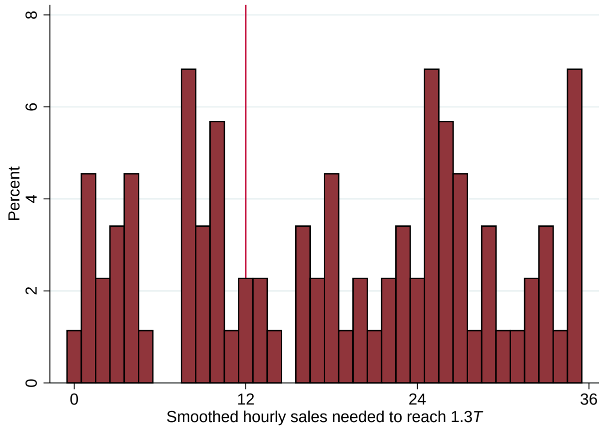
Figure C.6.1 shows simple frequency distributions of the number of sales hours needed to reach each of the three commission thresholds on the morning of all last days of the month, in all our teams. Since each store operates roughly 12 hours per day, effective advance planning should generate either excess mass or missing mass at that point (shown by the red vertical line). We see little evidence of either form of advance planning.

Figure C.6.1: Number of Hours Needed to Reach the Next Threshold

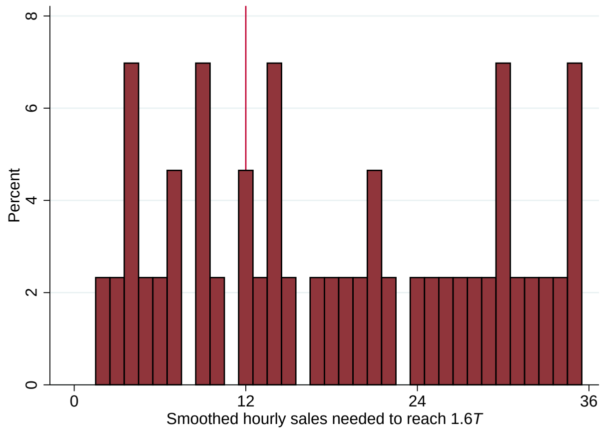
(a) T



(b) $1.3T$



(c) $1.6T$



Notes: This figure plots the distribution of number of hours (measured by smoothed hourly sales) needed to reach the next threshold (T, $1.3T$, or $1.6T$), on the morning of the last day of a month. The smoothed hourly sales is calculated by dividing the threshold sales amount by (number of days in the month $\times$ 12), as each store operates roughly 12 hours a day. The red vertical line at 12 indicates that currently it is approximately one day's work away to reach the next threshold.

C.6.2 Dropping Days that are Most Likely to be Affected by Advance Planning

Another way to assess the effects of advance planning on our estimation results is to remove days in which team output is most likely to be affected by advance planning from our estimation sample. Since advance planning becomes easier and easier to do as the end of the month approaches (after most of the random shocks to daily sales have been realized), this involves dropping non-last days that are later in the month. This exercise has already been performed in Appendix C.2, which shows that dropping either the last week or the last three weeks has little effect on our Table 3 estimates.

C.6.3 Implications of Advance Planning

Despite finding no evidence of advance planning, and despite the robustness of our results to dropping sample days that are most likely to be affected by advance planning, it is still prudent to explore what *would* be the implications of advance planning for our ability to distinguish between classical and other models of team behavior. In the rest of this section, we explain why forward-looking planning by teams –if it existed– would not be a threat to our main claim in the paper, which is that teams treat all three thresholds in their pay schedules as goals that yield psychological value.

To see this, consider first the case where inherited output on the last day of the month is randomly assigned. Then it is clear that our estimates are causal effects of the presence (and distance) of an attainable threshold on the last day on a team's output levels. If, in this context, we observed gaps in teams' output distributions around attainable 'pure' kinks (1.3T and 1.6T), that would be evidence in favor of the standard model. If instead we saw bunching at these kinks, that would be evidence against the standard model and in favor of our alternative, 'psychological' models.

Now consider how this reasoning changes if teams start planning their efforts several days before the end of the month. Teams that wish to avoid pure kinks (i.e. teams governed by the standard model) should try to position themselves so that, on the morning of the last day, they are in a good position to 'miss' these thresholds on the last day: a rational team will position itself to either go beyond a concave kink or not try to attain it. So classical, forward-looking teams are still predicted to have gaps in their final-day output distributions at 1.3T and 1.6T (perhaps even more so since they have more ways to optimize their outputs). The same reasoning applies to teams that treat 'kinks as goals': These teams should work to position themselves so that –on the morning of the last day– they are in a position to 'just attain' either 1.3T or 1.6T. Thus, evidence of bunching at 1.3T or 1.6T on the last day is still evidence in favor of a psychological model of goals.

In sum, our bunching-based test between the psychological model and classical models of team behavior is robust to the presence of advance planning by teams. This is because the models we are testing have the same (conflicting) predictions for the distribution of sales on last days, *whether or not* teams plan ahead.

D Appendix D: Heterogeneity

In this Appendix, we assess how our main results about output bunching vary with team and store characteristics, and with each team's recent sales and turnover history. To economize on the number of coefficients reported, this Appendix focuses on Table 4's coarser categorization of sales bins, and reports only the results for the bin containing the threshold and the two aggregated bins on either side of it.

D.1 Team Size and Expected Sales

In this section, we examine whether the amount of bunching varies with the size of the team. To that end, Table D.1.1 replicates Table 4 allowing for different effects of thresholds on the sales distribution in stores with two employees, three employees, or four or more employees. Notably, at both the target and the two kinks we find statistically significant bunching in teams of three and four-plus workers, but no bunching in teams of only two workers. Also, the estimated magnitude of bunching at the both the target and kinks increases strongly and monotonically with team size. These findings are confirmed graphically from the raw sales distributions shown in Figure D.1.1, which replicates Figure 5 by team size. At all thresholds, Figure D.1.1 shows more bunching in stores of at least 4 workers, compared to smaller stores.

As discussed in more detail in Appendices E.1.6 and E.3, the results in Table D.1.1 and Figure D.1.1 are strikingly at odds with the idea that reward misperception explains bunching (because larger teams should make 'better' decisions), and with the idea that team leaders use material leverage over subordinates to coerce them into bunching (because leverage over a single co-worker should be much greater than larger team.) Instead, the results suggest that thresholds might be functioning as coordination devices for team production.

In Table D.1.2, we disaggregate stores by using a different measure of size– their typical daily sales. To do so, we first calculate the typical daily sales output of a store-month by $TDS = T/D$, where T is a store's target in the current month, and D is the number of days in the month. Here, there are only small differences in daily sales between larger and smaller stores. Consistent with our team-coordination hypothesis, the number of members and not the dollar value of total team sales is the critical factor affecting the amount of bunching.

Table D.1.1: **Heterogeneity Examination – Team Size Effects**

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month	
	T	$1.3T$ & $1.6T$
Size=2:		
$K_{-5,-1}\times$ (Size=2)	-0.0127	0.0275
	(0.1408)	(0.1402)
$K_0\times$ (Size=2)	**-0.0866**	**0.0213**
	(0.0758)	**(0.0861)**
$K_{1,5}\times$ (Size=2)	0.0955	-0.0734
	(0.0960)	(0.0559)
Size=3:		
$K_{-5,-1}\times$ (Size=3)	-0.0622	-0.0619
	(0.0491)	(0.0417)
$K_0\times$ (Size=3)	**0.0965****	**0.1067****
	(0.0454)	**(0.0456)**
$K_{1,5}\times$ (Size=3)	0.0955*	0.0175
	(0.0545)	(0.0459)
Size$\geq$4:		
$K_{-5,-1}\times$ (Size$\geq$4)	-0.0862*	-0.0461
	(0.0502)	(0.0608)
$K_0\times$ (Size$\geq$4)	**0.1442*****	**0.2050****
	(0.0492)	**(0.0801)**
$K_{1,5}\times$ (Size$\geq$4)	0.1155*	-0.0339
	(0.0655)	(0.0604)
N	2697816	2697791

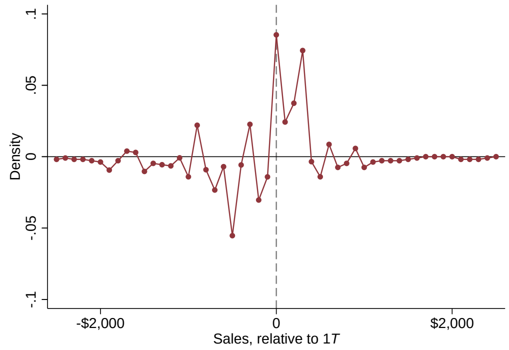
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This Table replicates Table 4, interacting regressors with indicators for team size. Observations that are five or more bins from the threshold are also included in the regression. To economize on degrees of freedom, the regressions for $1.3T$ and $1.6T$ are combined. Standard errors reported in parentheses are clustered at the store level. Coefficients are rescaled as in Table 4. Please see notes to Table 4 for additional details.

Figure D.1.1: Excess (or Deficient) Sales Relative to the threshold (By Store Size)

T

(a) Size≤ 3



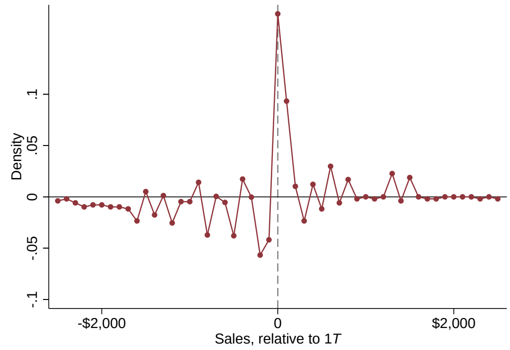
H_0: Equal density in Bin 0

t = 2.618; p-value = 0.009

H_0: Equal density in Bins -100 $\sim$ 100

t = 1.863; p-value = 0.063

(b) Size≥ 4



H_0: Equal density in Bin 0

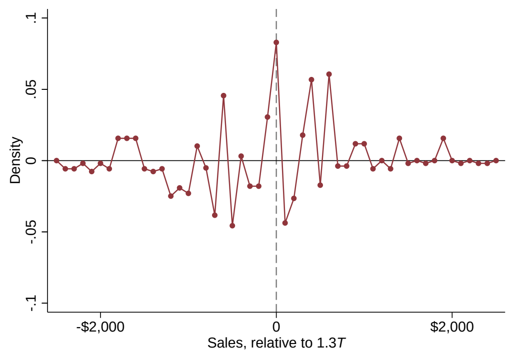
t = 4.605; p-value = 0.000

H_0: Equal density in Bins -100 $\sim$ 100

t = 3.720; p-value = 0.000

$1.3T$

(c) Size≤ 3



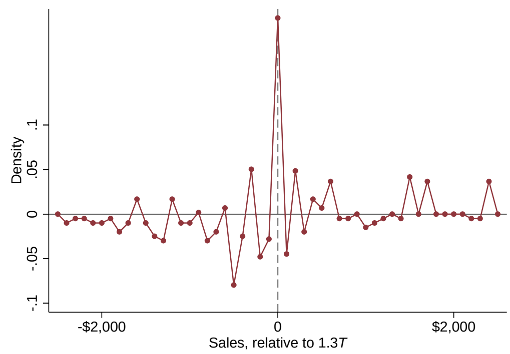
H_0: Equal density in Bin 0

t = 2.252; p-value = 0.025

H_0: Equal density in Bins -100 $\sim$ 100

t = 1.028; p-value = 0.305

(d) Size≥ 4



H_0: Equal density in Bin 0

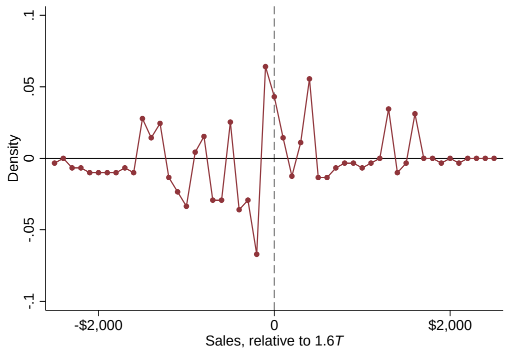
t = 4.666; p-value = 0.000

H_0: Equal density in Bins -100 $\sim$ 100

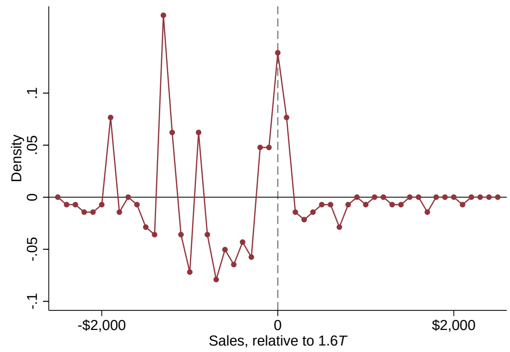
t = 1.798; p-value = 0.074

1.6T

(e) Size$\leq$ 3



(f) Size$\geq$ 4



H_0: Equal density in Bin 0

t = 1.002; p-value = 0.317

H_0: Equal density in Bins -100 $\sim$ 100

t = 1.870; p-value = 0.063

H_0: Equal density in Bin 0

t = 2.047; p-value = 0.043

H_0: Equal density in Bins -100 $\sim$ 100

t = 2.729; p-value = 0.007

Last days (thresholds within reach) Non-last days (thresholds within reach)

Notes: These figures plot the histogram of excess (or deficient) sales relative to the corresponding kinks, measured in bins of \$100, separately for last days and non-last days. For better visual display, only the range [-\$2,500, \$2,500] is presented. The figures are plotted for T, 1.3T, and 1.6T, separately. On the left-hand side, the sample is restricted to stores which have no more than 3 workers; and on the right-hand side, the sample is restricted to stores which have at least 4 workers. Please see Figure 5 for figures of the full sample.

Table D.1.2: Heterogeneity Examination – By Size of a Store's Typical Daily Sales

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month	
	T	$1.3T$ & $1.6T$
Low sales team months (TDS$\leq$500):		
$K_{-5,-1}\times$ i.(TDS$\leq$500)	-0.0550	-0.0494
	(0.0598)	(0.0531)
$K_0\times$ i.(TDS$\leq$500)	**0.0704**	**0.0987****
	(0.0473)	**(0.0467)**
$K_{1,5}\times$ i.(TDS$\leq$500)	0.0760	-0.0110
	(0.0515)	(0.0454)
High sales team months (TDS$>$500):		
$K_{-5,-1}\times$ i.(TDS$>$500)	-0.0553	-0.0223
	(0.0354)	(0.0384)
$K_0\times$ i.(TDS$>$500)	**0.1321*****	**0.1283****
	(0.0482)	**(0.0560)**
$K_{1,5}\times$ i.(TDS$>$500)	0.1115*	-0.0180
	(0.0605)	(0.0414)
N	2697816	2697791

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This Table replicates Table 4, interacting regressors with indicators of high or low typical store sales. Observations that are five or more bins from the threshold are also included in the regression. TDS is calculated for each store-month by TDS=T/D, where T is the store's monthly target, and D is the number of days in the current month. To economize on degrees of freedom, the regressions for $1.3T$ and $1.6T$ are combined. Coefficients are rescaled as in Table 4. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 4 for additional details.

D.2 Team Experience Effects

Table D.2.1 asks whether experienced sales teams are more or less likely to bunch at thresholds. As shown in Table D.2.1, the amount of bunching at both the target and the kinks increases with average firm tenure of the sales teams. As discussed in in more detail in Appendices E.1.6 and E.3, these results are inconsistent with the hypothesis that bunching at the kinks is caused by workers' misperceptions of Firm A's personnel policies.

Table D.2.1: Heterogeneity Examination – Firm Tenure Effects

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month	
	T	$1.3T$ & $1.6T$
$K_{-5,-1}\times$ c.Tenure	-0.0062	-0.0106
	(0.0164)	(0.0068)
$K_0\times$ c.Tenure	**0.0128****	**0.0190****
	(0.0058)	**(0.0090)**
$K_{1,5}\times$ c.Tenure	0.0140	-0.0009
	(0.0161)	(0.0082)
N	2697816	2697791

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This Table replicates Table 4. In addition to variables identifying bins on either side of the threshold as in Table 4, each regression adds regressors that interact the threshold-containing bin with a continuous measure of the team's mean tenure with Firm A. To economize on degrees of freedom, the regressions for $1.3T$ and $1.6T$ are combined. Like Table 4, all regressions include variables identifying bins on either side of the threshold. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 4 for additional details.

D.3 Promotion Prospects

Table D.3.1 examines whether bunching is more pronounced in teams whose managers can possibly receive promotions into managerial positions in the head office. If a store manager believes that 'just' crossing a threshold discretely increases her promotion chances into managerial positions, she might try to motivate her teammates to do so. As managerial positions are only offered at the head office, we use the location of the retail store to proxy for the manager's promotion chances. As shown in Table D.3.1, promotion prospects do not significantly affect the likelihood of bunching in our context.

Table D.3.1: Heterogeneity Examination – Promotion Prospects

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month	
	T	$1.3T$ & $1.6T$
$K_{-5,-1}\times$ i.(Near HQ)	0.1728**	0.0044
	(0.1041)	(0.0674)
$K_0\times$i.(Near HQ)	**-0.0804**	**0.0142**
	(0.0667)	**(0.0681)**
$K_{1,5}\times$ i.(Near HQ)	0.0385	0.0160
	(0.0800)	(0.0689)
N	2697816	2697791

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This Table replicates Table 4. In addition to variables identifying bins on either side of the threshold as in Table 4, each regression adds regressors that interact the corresponding bins with an indicator for a store's location. Near HQ is an indicator variable, taking a value of 1 if the store locates in the same city where the head office locates. As managerial positions are only offered at the head office, it is a proxy for higher promotion chances. To economize on degrees of freedom, the regressions for $1.3T$ and $1.6T$ are combined. Coefficients are rescaled as in Table 4. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 4 for additional details.

D.4 Turnover Events

Table D.4.1 estimates whether the amount of bunching at thresholds responds to recent turnover events at the store. For example, disruptions to team membership might make it harder for teams to agree on a bunching strategy. We find that the likelihood of bunching at both the target and kinks is not significantly different in months when a turnover has occurred, or in months following a turnover event, compared to other months.

Table D.4.1: Heterogeneity Examination – Turnover Events

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month			
	T	$1.3T$ & $1.6T$	T	$1.3T$ & $1.6T$
$K_{-5,-1}\times$ Turnover$_{im}$	-0.0725	-0.0773		
	(0.0969)	(0.1010)		
$\boldsymbol{K_0}\times$ **Turnover**$_{im}$	**0.0861**	**0.1267**		
	(0.1032)	**(0.1086)**		
$K_{1,5}\times$ Turnover$_{im}$	-0.1310	0.0190		
	(0.0935)	(0.1063)		
$K_{-5,-1}\times$ Turnover$_{im2}$			-0.1219	-0.0835
			(0.0989)	(0.0904)
$\boldsymbol{K_0}\times$ **Turnover**$_{im2}$			**0.0843**	**0.0782**
			(0.0887)	**(0.0905)**
$K_{1,5}\times$ Turnover$_{im2}$			-0.0430	0.0185
			(0.0930)	(0.0884)
N	2697816	2697791	2697816	2697791

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This Table replicates Table 4. In addition to variables identifying bins on either side of the threshold as in Table 4, each regression adds regressors that interact the corresponding bins with indicators for recent turnover on the sales team. Turnover$_{i,m}$ is an indicator variable, taking a value of 1 if there is turnover event during the current month m at store i. Turnover$_{i,m2}$ is an indicator variable, identifying month m if there is turnover event at store i, and identifying the month $m+1$ following the turnover. To economize on degrees of freedom, the regressions for $1.3T$ and $1.6T$ are combined. Coefficients are rescaled as in Table 4. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 4 for additional details.

D.5 Host Institution Type

Firm A's stores are located in two different types of host institutions –departments and shopping malls– which interact with the stores in slightly different ways. For example, department stores collect part of the store's monthly revenues, while shopping malls collect a fixed amount of monthly rent. Also, other department store employees provide occasional assistance to Firm A's salespeople in ways that are absent in shopping malls. To see if these small institutional differences affect the amount of sales bunching at Firm A's commission kinks, Table D.5.1 interacts the threshold and surrounding bins with an indicator for department store host institutions. It shows that bunching behaviors are not associated with the type of host institution.

Table D.5.1: Heterogeneity Examination – Department Stores vs. Shopping Malls

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month	
	T	$1.3T$ & $1.6T$
$K_{-5,-1}\times$ i.DepStore	0.0344	0.0098
	(0.0229)	(0.0247)
$\boldsymbol{K_0}\times$ **i.DepStore**	**-0.0733**	**0.0015**
	(0.0705)	**(0.0903)**
$K_{1,5}\times$ i.DepStore	0.0086	-0.0038
	(0.0177)	(0.0170)
N	2697816	2697791

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This Table replicates Table 4. In addition to variables identifying bins on either side of the threshold as in Table 4, each regression adds regressors that interact the corresponding bins with indicators for host institutions being department stores, rather than shopping malls. (Department stores collect part of the store monthly revenues, while shopping malls collect a fixed amount of monthly rent.) To economize on degrees of freedom, the regressions for $1.3T$ and $1.6T$ are combined. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 4 for additional details.

D.6 Store Sales Volatility

If bunching at commission kinks is facilitated by advance planning before the end of the month, we should see less bunching in stores that experience high day-to-day sales volatility. This is because advance planning is more difficult, and less effective when future sales are hard to forecast. To see if this is the case, Table D.6.1 uses each store's day-to-day sales variance in 2015 - the year prior to our sample year- to measure sales volatility. As shown in Table D.6.1, sales volatility at a store does not affect the amount of bunching. Consistent with other evidence summarized in Appendix C.6, this suggests that advance planning may not be an important feature of team behavior at Firm A.

Table D.6.1: Heterogeneity Examination – By Store Sales Volatility

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month	
	T	$1.3T$ & $1.6T$
$K_{-5,-1}\times$ i.VolStore	-0.0196	0.0203
	(0.0208)	(0.0238)
$K_0\times$ i.VolStore	**0.0833**	**-0.0282**
	(0.0668)	**(0.0875)**
$K_{1,5}\times$ i.VolStore	0.0159	0.0196
	(0.0153)	(0.0159)
N	2697816	2697791

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This Table replicates Table 4. In addition to variables identifying bins on either side of the threshold as in Table 4, each regression adds regressors that interact the corresponding bins with indicators for high store sales volatility. Store sales volatility is measured using variance in log(sales) in 2015, and stores with above-median volatility are indicated as the volatile stores. To economize on degrees of freedom, the regressions for $1.3T$ and $1.6T$ are combined. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 4 for additional details.

D.7 Previous Threshold Attainment

According to some models of loss aversion, teams that have recently attained a particular threshold may become 'attached' to it, and begin to treat it as a reference point that induces loss aversion (DellaVigna et al., 2017). To see if this is the case, we calculate the number of times a team reached each threshold in the first three months of 2016, and then interact K_0 with this variable in regressions estimated on the remaining nine months of 2016.[4] As shown in Table D.6.1, there is no evidence that bunching behavior is associated with the number of times a threshold was recently attained. In combination with other results –such as the increase in sales density to the right of pure kinks– we see this as evidence against a loss-aversion-based explanation of the unexpected sales bunching we document in this paper.

Table D.7.1: Heterogeneity Examination - By Previous Threshold Attainment

	Dependent variable: Daily Sales (in US$)	
	T	$1.3T$ & $1.6T$
$K_{-5,-1}\times$ c.# Attained in M1-M3	0.0673	-0.0084
	(0.0917)	(0.0527)
$K_0\times$ c.# Attained in M1-M3	**-0.0089**	**-0.0012**
	(0.0772)	**(0.0630)**
$K_{1,5}\times$ c.# Attained in M1-M3	-0.0185	0.0719
	(0.0745)	(0.0434)
Store$\times$ DOW	Yes	Yes
Holiday FE	Yes	Yes
N	2018916	2018200

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This Table replicates Table 4. In addition to variables identifying bins on either side of the threshold as in Table 4, each regression adds regressors that interact the corresponding bins with the variable # Attained in M1-M3. The variable # Attained in M1-M3 calculates how many times a team has reached the corresponding threshold in the first three months. In column (2), this variable calculates how many times a team has reached $1.3T$. The regression is estimated for observations for the rest of the year in months 4-12. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 4 for additional details.

[4] We cannot use threshold attainment in 2015, because we do not have that information.

E Appendix E– Alternative Explanations

This Appendix supplements Section 6.4 of the paper, which assesses alternative explanations of the bunching at the kinks of Firm A's commission schedule. Specifically, Appendix E.1 provides additional details relevant to Section 6.4.1's analysis of unobserved discrete material rewards (UDMRs). Appendix E.2 provides Table E.2.1, which is relevant to Section 6.4.4's analysis of the intertemporal sales shifting hypothesis.

E.1 Unobserved Discrete Material Rewards (UDMRs)

In general, an unobserved discrete material reward is any unobserved aspect of a worker's pay, future prospects, or utility that changes discretely when her team's monthly sales pass one of the three thresholds (T, $1.3T$, of $1.6T$). Importantly, UDMRs do not include connections between these outcomes and *continuous* indicators of a team's sales performance, including, for example, practices that link promotions or retention to a continuous measure of a team's sales *relative* to its targets. Only outcomes that change discretely when teams pass their commission thresholds can cause the bunching we are trying to explain.

We start our analysis of UDMRs by reminding readers that Firm A views its sales teams more as independent, self-managing units than as the bottom rung of a corporate hierarchy. Indeed, Firm A's management prefers to minimize its involvement in the daily and monthly operations of the teams, relying instead on strong, team-based commissions to encourage co-operation, problem-solving, and peer oversight *within* each of over 100 teams in widely scattered locations. Consistent with this view, Firm A collects little or no information about the performance of individual workers within the teams: recall that no individual sales data are collected. Also, as further documented below, promotions out of the teams into management are nonexistent for team members, and very rare for team leaders. Overall, career incentives and monitoring of individual sales workers are both minimal, with work incentives relying heavily on the team-based commission.

Next, we consider four channels via which UDMRs could conceivably be tied to team kink attainment: base salaries, layoffs, team member promotions, and team leader promotions.

E.1.1 Base Salaries

Base salaries can easily be eliminated as a channel for UDMRs because – as described in the paper – individual workers' base salaries are determined by a formula that depends only on their tenure with the firm and their store's location.

E.1.2 Layoffs and Dismissals

While some firms undoubtedly dismiss or lay off individual workers based on their job performance, management at Firm A does not dismiss individual sales workers for poor performance, or lay off individual sales workers for business reasons. As already noted, Firm

A has no data on individual workers' sales, so it cannot single out low-sales workers for dismissal. Firm A also has a policy of not intervening in within-team disciplinary issues, preferring instead to incentivize the teams to solve such issues independently with strong, team-based cash incentives. In addition, Firm A has a long-standing, public, no-layoff policy for its sales workers. The firm has honored this policy consistently since at least 2008 (Kuhn and Yu, 2021). In fact, even when stores are closed, sales positions at nearby locations are offered to all the affected employees. Consistent with this policy, there are also no recorded layoffs or dismissals during our sample period.[5] In sum, management-initiated dismissals or layoffs are simply not a concern for sales workers at Firm A.

The one way in which under-performing individual sales workers might be induced to leave Firm A is through pressure from their teammates; in fact this is a natural (and perhaps intended) consequence of Firm A's team-based commission scheme (Kuhn and Yu, 2021). While we suspect that peer pressure plays an important role in maintaining team discipline at Firm A, we cannot think of any 'classical' reason why teams would use this type of peer pressure to encourage team members to bunch at commission kinks. Indeed, teams of 'classical' workers should use any such leverage to *avoid* bunching. For this reason, teams' use of peer pressure to discipline members does not invalidate our tests between the classical and alternative models of team behavior.

E.1.3 Promotions of Team Members

Firm A does not promote team members (who are not leaders) to jobs in the head office. Promotions of members to team *leader* positions do, of course, occur, but only when only when the current leader leaves the store. The extrinsic rewards to becoming a team leader are confined to a small increase in base salary **(about X, or Y percent of mean total monthly pay)** to compensate for the increased administrative burden. As described in section E.1.5 below, the prospects of promotion beyond the team leader level are also very low. New team leaders are selected by the firm's regional managers, usually on the basis of seniority.

Most importantly, since one (and only one) member is promoted to leader when a departure occurs, there is no way for Firm A to link the probability that a **representative** member is promoted to leadership to *any* measure of team performance. That probability is always $1/(N-1)$ times the probability that the team leader departs, where N is the size of the team (including the leader). In sum, there is no plausible hidden link between a team member's promotion prospects and whether (or how frequently) her team's sales pass the thresholds in Firm A's team commission scheme.

E.1.4 Promotions of Team Leaders

In our assessment, the most likely way that UDMRs could be attached to a team's attainment of discrete sales thresholds is via team *leaders'* promotion prospects. For this to occur, however, all of the following conditions would need to be satisfied:

[5]Notice also that the dominant role of commissions in total pay at Firm A will automatically incentivize workers in ailing stores to quit.

1. Promotions of team leaders to head office should be reasonably frequent relative to Firm A's high turnover rates, and should be at least somewhat lucrative.

2. Team leaders' promotion prospects must depend, discontinuously, on whether their team's monthly sales pass each of the three thresholds in the commission schedule.

3. Team leaders must have some material leverage over the other team members that can be used to incentivize the members to expend effort to raise the leader's promotion prospects.

With respect to condition 1, we note that promotions of team leaders to the head office are rare: in fact only one team leader was promoted to head office in 2016. One reason these promotions are so rare because head office positions are primarily filled via outside hires or promotions within the head office. Another is that, among team leaders, promotions are usually only offered to the leaders who live near Firm A's head office: More-distant workers would not find such jobs attractive, given the high commuting costs and the low salary increment associated with them.[6]

Turning to condition 2, we asked management which criteria they use when making these (rare) promotion decisions. They indicated that they do not consider the sales performance of the managers' teams, in either continuous or discretized forms. Instead, because head office jobs are fundamentally different from sales jobs (Benson et al., 2019), Firm A focuses on the manager's educational background and communication skills. With respect to discretized team performance measures (such as the number of times a team passed any one of the three performance thresholds) Firm A further clarified that its internal personnel and sales accounts contain only the continuous version of each employee's monthly pay and each team's monthly sales. Thus, while counts of threshold attainment could of be constructed by merging these records with each store's monthly targets, Firm A's record-keeping practices are support its claims that indicators of team threshold attainment, *per se*, are not used in *any* of its management or personnel decisions (other than to calculate each worker's monthly commission pay).

Finally, turning to condition 3 (leaders' leverage over their team-mates), we note that team leaders have no discretion over layoffs (which are not used by Firm A in any case), pay, or promotions of the other team members. While it is possible that leaders have some additional, unobserved leverage, we note that our team size heterogeneity analysis in Appendix D.1 is inconsistent with leaders using unobserved leverage to encourage members to bunch at commission kinks. Specifically, any managerial leverage over team members should be more effective in small teams than in larger ones. (For example, the manager of a two-person team can reward or punish half the team by rewarding or punishing a single worker). Yet we find the strongest bunching at kinks in the *largest* teams.

Summing up our description of Firm A's *actual* personnel policies, we first reiterate that Firm A firmly denies attaching UDMRs to the thresholds in its team commission schedule. Next, based on

[6]The annual pay of a head office job is at about the 65th percentile of store managers', though with higher commute costs even for workers who live close enough to commute.

a detailed analysis of Firm A's actual personnel policies, and on a case-by-base analysis of possible mechanisms, we cannot see any realistic ways in which Firm A might be *inadvertently* linking outcomes that workers care about to team threshold attainment *per se*.[7]

E.1.5 Workers' Perceptions of Firm A's Personnel Policies

Despite what Firm A says and does, could workers mistakenly *believe* that they will receive some discrete material reward if their team exceeds the sales thresholds in their commission schedule? While misperceptions by individual workers are always possible, we start our discussion of this question by noting that misperceived personnel policies are a threat to our bunching results only if they are *shared* by the members of a work team. Shared mistakes seem less likely, in part because informed group members can correct others' mistakes (Charness and Sutter, 2012).

Next, we describe three results from our heterogeneity analyses that are inconsistent with the idea that workers falsely believe that discrete material rewards (other than the jump in the *marginal* commission rate) are attached to Firm A's commission kinks. First, if bunching is driven by misperceived rewards, it should be mitigated by feedback and experience. Despite this, we find that bunching at all three commission thresholds increases with team experience. If anything, our teams learn *to* bunch, not to avoid bunching. Second, if managers falsely believe that hitting targets raises their promotion chances, there should be more sales bunching in stores located close to Firm A's head office, compared to other stores. There is no difference, however.[8]

Third, we also find that bunching is much more pronounced in larger than smaller teams; this is inconsistent with misperceived UDMRs for two reasons. One, already noted, is that larger teams offer more opportunities for informed group members to correct others' misperceptions. The other relates to the well known $1/N$ problem in teams. Specifically, suppose that – contrary to Firm A's claims – Firm A rewards individual members of teams that pass thresholds. In that case, bunching should be less pronounced in larger teams, because individual workers' sales efforts have a much smaller influence on the probability of reaching a team threshold in larger teams.

We conclude this section by asking whether workers might have any *reason to doubt* Firm A's claims that the sales thresholds are only used to calculate commission rates. Workers should see such claims as more credible if Firm A's pay, retention and promotion policy choices are well understood by workers, and are seen to fit together as part of profit-maximizing personnel system. In our assessment, this is the case, and it is well understood by Firm A's sales workers.

Specifically, Firm A's practice of using *only* team-based performance measures, abstaining from

[7]Additional material outcomes that could, in principle, be linked to a team's threshold attainment include store-level outcomes like avoiding closure, or qualifying for an expansion in team size. Some of these linkages are made less likely by other institutional details (for example, all workers in closed stores are offered jobs in other stores; individual commission rates are lower in larger teams, in order to keep total pay about the same). More importantly, all these potential linkages face a now-familiar challenge: rational firms should base any such linkages on the *continuous* performance measures in the company's files, not on discretized performance measures (which are not in those files.)

[8]Notice that this result rules out both worker misperceptions about how managers are promoted, and any actual connections that might exist between teams' threshold attainment and manager promotions, despite Firm A's insistence to the contrary.

dismissals and layoffs, and not intervening in the day-to-day functioning of the sales teams is well adapted to Firm A's sales environment. This system is intended to encourage worker cooperation within each of over 100 teams scattered across different locations, and to encourage those teams to function independently with little oversight from management (Kuhn and Yu, 2021). Essentially, Firm A's sales teams are not the bottom rung of a corporate hierarchy; they are independent groups of high-turnover workers who interact almost exclusively with their team-mates, and who are motivated almost exclusively by a commission schedule that strongly links each member's pay to their team's total output.

E.2 Testing for Sales Shifting

Table E.2.1 tests for the type of time-shifting that a 'non-classical' team would use if it wanted to 'just' reach a threshold on the last day of a month. Such teams would have an incentive to shift sales from the beginning of the next month into the last day of the current month when a threshold is within reach. As shown in Table E.2.1, sales early in a month are *not* lower when the team had an attainable threshold on the last day of the previous month. Consistent with the fact that sales shifting is extremely difficult at Firm A, we conclude that the bunching detected by our regressions in Tables 3, 4 and 6 represents true increases in sales effort, not strategic changes in the timing of sales.

Table E.2.1: **Testing for Sales Shifting**

	Dependent variable: Daily Sales (in US$)		
	T	$1.3T$	$1.6T$
Specification 1:			
T-day +1	173.4	564.2	447.6^{*}
	(185.8)	(345.0)	(230.7)
T-day +2	-37.5	-27.4	-29.0
	(86.7)	(103.1)	(159.4)
T-day +3	-245.6^{**}	-181.2^{*}	-41.4
	(94.7)	(97.1)	(111.0)
Specification 2:			
T-day +1 $\sim$ +3	-36.6	119.7	126.6
	(95.9)	(140.4)	(121.5)
Store$\times$ DOW	Yes	Yes	Yes
Holiday FE	Yes	Yes	Yes
N	34863	34863	34863

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Each observation in this table is a daily observation. The dependent variable is the daily sales amount. In specification 1, the independent variables are binary, taking a value of 1 if the focal day is 1, 2, or 3 days after the last day when a corresponding threshold is within reach. In specification 2, the variable of interest is binary, taking a value of 1 if the focal day is within three days following a last day when the corresponding threshold is within reach. In all regressions, store $\times$ day-of-the-week fixed effects and holiday fixed effects are controlled. Standard errors reported in parentheses are clustered at the store level.

References

Benson, A., Li, D., and Shue, K. (2019). Promotions and the peter principle. *The Quarterly Journal of Economics*, 134(4):2085–2134.

Card, D., Mas, A., Moretti, E., and Saez, E. (2012). Inequality at work: The effect of peer salaries on job satisfaction. *American Economic Review*, 102(6):2981–3003.

Charness, G. and Sutter, M. (2012). Groups make better self-interested decisions. *Journal of Economic Perspectives*, 26(3):157–76.

DellaVigna, S., Lindner, A., Reizer, B., and Schmieder, J. F. (2017). Reference-dependent job search: Evidence from hungary. *The Quarterly Journal of Economics*, 132(4):1969–2018.

Kőszegi, B. and Rabin, M. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics*, 121(4):1133–1165.

Kuhn, P. J. and Yu, L. (2021). How costly is turnover? evidence from retail. *Journal of Labor Economics*.

MacKinnon, J. G. and Webb, M. D. (2020). Randomization inference for difference-in-differences with few treated clusters. *Journal of Econometrics*, 218(2):435–450.

Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy*, 2(3):180–212.