

The Effects of Quota Frequency: Sales Performance and Product Focus

Doug J. Chung, Harvard University

Das Narayandas, Harvard University

Dongkyu Chang, City University of Hong Kong*

February 2020

Forthcoming at *Management Science*

Abstract

This study investigates the comprehensive and multidimensional effects of quota (goal) frequency on sales force performance. The study provides a theory of salespeople's behavior—aggregate effort and the product type focus—in response to the temporal length of a sales-quota cycle. The theory includes many realistic elements, such as salespeople's multi-dimensional effort, heterogeneity in ability, product focus, and forward-looking behavior. We test the theory through a field experiment, varying the sales compensation structure of a major retail chain in Sweden. Consistent with the developed theory, shifting to a temporally frequent quota structure leads to an increase in sales performance for low-performing salespeople by preventing them from giving up in later periods within a quota-evaluation cycle, but to a decrease in sales performance for high-performing salespeople. With quotas set over short time horizons, the high-performing salespeople focus mainly on low-ticket products, resulting in a decrease in both sales volume and the sale of high-ticket products, thus reducing the firm's profits.

Key words: sales force compensation, field experiment, quota, goal, product focus.

* Doug J. Chung is the MBA Class of 1962 Associate Professor of Business Administration (dchung@hbs.edu) and Das Narayandas is the Edsel Bryant Ford Professor of Business Administration (mnarayandas@hbs.edu) at Harvard University. Dongkyu Chang is an Assistant Professor of Economics at the City University of Hong Kong.

The authors thank the department editor, associate editor, and three anonymous reviewers for their constructive feedback during the review process. The authors also thank Brian Hall, Thomas Keifer, Martin Knutson, Gary Lilien, Joel Rönneman, K. Sudhir, and the seminar participants at the California State University Long Beach, Harvard University, Sungkyunkwan University, University of North Carolina at Charlotte as well as the conference participants at the 2017 Consortium for Operational Excellence in Retailing (COER), 2017 INFORMS Marketing Science, and the 2018 Yale Customer Insights conferences for their comments and suggestions. The authors also thank the anonymous firm for cooperating in this study.

1. Introduction

Incentives are ubiquitous, especially in a capitalistic free-market economy. They are believed to provide one of the primary motivations for people to work, particularly in the domain of personal selling. Despite recent advances in sophisticated marketing techniques using big data and artificial intelligence to persuade customers and encourage purchases, personal selling still remains a significant (and, in most industries, the only) function in firm–customer interactions. According to the U.S. Bureau of Labor Statistics, 15 million people, or about 10% of the entire U.S. labor force, are employed in personal selling (U.S. Department of Labor, 2018). U.S. firms spend over \$800 billion annually on personal selling, an amount that is more than three times greater than the total expenditures on media and digital advertising (\$208 billion) (Zoltners, Sinha, & Lorimer, 2013). A large proportion of spending on personal selling is used to incentivize and motivate salespeople to exert greater selling effort. With so many people and resources at stake, the design of the sales compensation system is of great strategic importance to firms.

A sales compensation system typically consists of a fixed salary plus variable pay, conditional on meeting a sales quota (i.e., achieving a certain threshold of performance). Firms commonly use quotas; in fact, about three quarters of U.S. firms use some form of quotas (Joseph & Kalwani, 1998). **Figure 1** shows illustrative examples of several quota-based compensation plans. Firms typically use quotas as achievement goals to evaluate performance and determine whether a salesperson has had a successful period (e.g., month or year). But how should a sales manager design a quota-based compensation plan? This study specifically attempts to answer, through theoretical illustration and empirical evidence via a field experiment, the following questions: What is the appropriate frequency of quotas? That is, at what intervals should quotas be set, and how often should they be evaluated? Would frequent quotas either increase or decrease sales performance?¹ If so, which types of salespeople would be affected? Does a salesperson’s quality of effort falter with frequent quotas?

¹ “Frequent quotas,” “frequent quota plans” or “short quota cycles” refers to quota-based plans that have more evaluation and payment periods than less-frequent quota plans with long quota cycles. In this study’s empirical context, a daily-quota plan represents a frequent quota plan, and a monthly-quota plan represents a less-frequent quota plan.

Would there be a change in a salesperson’s product focus if the frequency of sales quotas were changed?

To answer these questions, we first develop a theoretical model of salespeople’s behavior regarding different quota frequency. The theoretical model considers that salespeople’s actions are multi-dimensional and that their selling abilities differ with regard to the product type (low-ticket versus high-ticket). A theoretical proof shows that it is in the firm’s best interest to provide the most-frequent quota-evaluation cycles for salespeople who have limited ability to sell high-ticket products. In contrast, it may be optimal for the firm to have less-frequent quota cycles for salespeople who are better at selling high-ticket products. The firm wants salespeople to focus their efforts on products that they are effective at selling—that is, salespeople with the ability to sell high-ticket products should focus on selling such products. However, because the incentives of the firm and the salespeople are misaligned, there exists a distortion in the salespeople’s actions. Consequently, with frequent quotas, salespeople that are a better fit for selling high-ticket products alter their behavior to focus on selling a greater number of low-ticket products, thus changing the type of products sold and potentially decreasing the firm’s aggregate sales and profits.

To validate the theoretical claims and empirically examine the effects of quota frequency on sales performance, we conducted a field experiment with full-time sales employees of a major Swedish retail chain. The firm implemented an intervention in which, holding everything else constant, it changed the sales compensation scheme from a monthly- to a daily-quota plan, with a control group of several stores whose salespeople did not encounter a change in compensation structure. The variation in performance between the salespeople who experienced the change and those who did not allows us to account for any seasonal and other exogenous fluctuations in order to analyze, as cleanly as possible, the pure causal effect of quota frequency on various dimensions of sales performance.

Overall, this study uniquely contributes to the literature in several ways. First, the study provides a theoretical model of the effect of quota frequency on salespeople’s behavior. The model includes many realistic elements, such as multi-dimensional effort (with regard to the product focus) and forward-looking behavior; combined, these predict not only aggregate sales outcomes but also

changes in the types of products sold. Second, the study validates the theoretical predictions using the results of a large-scale field experiment involving more than 300 full-time salespeople of a major retail chain. Third, to the best of our knowledge, this study is one of the first to directly examine quota frequency in a sales force setting.² Fourth, the study examines the effect of quota frequency not only on sales revenue, but also on various other dimensions of performance, such as the quality and type of effort, by monitoring and measuring product returns and changes in product focus. Finally, the study considers the heterogeneous effects of quota frequency across different types of salespeople.

Substantively, through theoretical illustration and the empirical results of the field experiment, the study's results show that a change from a less-frequent (monthly) to a frequent (daily) quota plan increases sales performance, but mainly for low-performing salespeople. Because every day is a fresh start under the daily plan, low-performing salespeople's motivations remained intact throughout the month, whereas under the monthly plan, these salespeople gave up if they were too far from meeting quota in the later days of the month. In contrast, there were negative effects of quota frequency on high-performing³ salespeople, as they distorted their efforts from high-ticket to low-ticket products, resulting in a change in the types of products sold and a decrease in aggregate sales, thus reducing the firm's profits. Surprisingly, in the daily-quota plan, even though salespeople were not penalized for returned merchandise, they did not over-aggressively sell, which could have increased product returns.

The rest of the paper is organized as follows. Section 2 summarizes the related literature, and Section 3 presents the theoretical model. Section 4 explains the institutional details of the firm and the field experiment design. Section 5 presents the empirical model-specification, and Section 6 discusses the results. Section 7 concludes.

2. Literature Review

² Chung, Steenburgh, and Sudhir (2014) explore the concept of quota frequency. However, their analysis is based on counterfactual simulations using estimates from their structural model and is not inferred directly from the data.

³ We define high-performing salespeople as those who are effective at selling high-ticket products.

Despite the ubiquitous use of sales quotas, academics have remained skeptical about their effectiveness. In the economics literature, assuming rational agents, there are two primary arguments against the use of sales quotas. First, the discrete and nonlinear nature of quotas commonly pushes salespeople to less powerful areas of incentives (Holmstrom & Milgrom, 1987; Lal & Srinivasan, 1993). That is, the motivating effects of achieving a sales quota diminish when a salesperson either has already surpassed the quota or is too far from achieving it—a salesperson who has no hope of meeting quota is likely to give up. Second, in a B2B environment (less so in a B2C retail environment such as our empirical setting), sales quotas may provoke salespeople merely to manipulate the timing of sales so that the quotas have no additional effect on performance (Oyer, 1998). For example, if a salesperson has already met quota, instead of booking realized sales in the current period, he or she can simply ‘push’ sales into the future to count toward the next period’s quota. Relatedly, a salesperson who has not met quota can ‘pull’ sales from the future (i.e., book scheduled future sales in the current period) to achieve quota. Increasing the temporal frequency of quotas (e.g., from monthly to daily) would make a quota-based plan similar to a linear commission plan. Hence, according to the above two arguments, frequent quotas would provide better constant motivation to salespeople, regardless of past cumulative sales, and would mitigate the timing manipulation of sales.

There is a vast literature in psychology on goals—such as sales quotas—and their effect on motivation (for an extensive survey, see Latham & Locke (1991)). The discussion of this literature is in two parts—that of goals, in general, and that of subgoals.

The “goal-gradient hypothesis” (Hull, 1932; Hull, 1938) postulates that people become more motivated—the goal gradient gets steeper—as their (perceived) progress nears a goal (Cheema & Bagchi, 2011; Kivetz, Urminsky, & Zheng, 2006; Nunes & Drèze, 2006). A goal, by definition, is a key reference point with regard to a focused activity (Heath, Larrick, & Wu, 1999). Hence, the characteristics of the value of attaining a goal can be similar to those of the value function in Prospect Theory (Kahneman & Tversky, 1979)—losses loom larger than gains, diminishing sensitivity from the origin: concave in gains and convex in losses. Therefore, as a person gets closer to a goal, his or her marginal motivation to achieve the goal becomes higher (steeper goal gradient).

Because sensitivity diminishes as one is further from a goal, a person would be less motivated in the initial stages; this is known as the ‘starting problem’ (Heath, Larrick, & Wu, 1999). Splitting a grand goal into multiple smaller subgoals will move a person relatively closer to the origin (reference point), mitigating the starting problem (i.e., sensitivity and, thus, motivation increases). However, there can also be a negative motivational effect of multiple subgoals compared to one grand goal.⁴ Although motivation increases as one nears a subgoal, it will substantially decrease (flat goal gradient) once the subgoal is achieved. That is, people will become complacent and reduce effort after achieving subgoals (Heath, Larrick, & Wu, 1999). Furthermore, attainment of subgoals may liberate or permit a person to pursue other goals (Amir & Ariely, 2008; Fishbach & Dhar, 2005; Fishbach, Dhar & Zhang, 2006).

Although there are both positive and negative motivating effects with regard to attaining subgoals, studies have found that the positive effects outweigh the negative effects—for example, Gal and McShane (2012) in the domain of debt management and Zhang and Gao (2016) in that of reward programs. Relatedly, in the education literature, researchers have found that frequent testing results in better performance outcomes for students (for an extensive survey, see Bangert-Drowns, Kulik, & Kulik (1991)).

The above survey of the relevant literature suggests that, in this study’s context, having frequent quotas (e.g., by splitting a monthly quota into many smaller daily quotas) should lead to an increase in salespeople’s motivation and, thus, enhance their sales performance. There are two primary reasons. First, because a daily-quota plan gives salespeople a fresh start each day, it should help them maintain high motivation throughout the month. For example, under a monthly-quota plan, a salesperson who experiences bad luck earlier in the month may decide to give up later in the month because there is no chance of meeting or exceeding the firm’s quota. This would not be the

⁴ It is worthwhile to distinguish the role of goals from the role of incentive structure as a whole. In this study’s context, a quota would serve as a goal, whereas the quota-compensation scheme would serve as the incentive structure. Aside from the negative motivating effect of post-goal achievement, negative consequences of incentives per se may also exist. For example, if a firm provides an incentive but then takes it away, an agent may decrease performance compared to the baseline (when the firm did not provide incentives) because of a decline in the agent’s intrinsic motivation (Lepper, Greene, & Nisbett, 1973; Chung & Narayandas, 2017).

case under the daily-quota plan, as every day would present the salesperson with a new chance to succeed. Second, the daily-quota plan would tap into a salesperson’s motivation more often, and, thus, there would be more instances of steeper goal gradient.

However, as Darmon (1997) indicates, to motivate salespeople to achieve objectives, quotas should be challenging. Splitting a grand quota into multiple finer (thus frequent) quotas would be the same as replacing one challenging quota with many less-challenging quotas. Using the principal-agent theory framework, Kim (1997) and Oyer (2000) illustrate that, under specific assumptions, a non-linear discrete quota-bonus compensation system (i.e., a long quota-cycle plan in our domain) can be optimal for the firm. Furthermore, the flexibility to intertemporally allocate effort across multiple periods may make the monthly plan more effective, as a change to a daily-quota plan may merely provoke income targeting within a particular day (Camerer et al., 1997).⁵ In addition to the abovementioned arguments, the daily-quota plan could potentially increase anxiety and stress among salespeople, as they may worry day in and day out about meeting quota, resulting in demotivation. Also, as we witnessed with Sears in the 1990s, Marsh in the 2000s, and more recently with Wells Fargo, there can also be negative effects of an overly aggressive incentive compensation system (Zoltners, Lorimer, & Sinha, 2016), leading to unethical behavior and fraud (Schweitzer, Ordóñez, & Douma, 2004). In our context, the daily-quota plan may induce salespeople to become overly aggressive by selling products that customers would return later, hurting the firm in the long run.

The literature review, thus far, has discussed the positive and negative effects of a frequent-quota compensation plan on aggregate sales, assuming unidimensional effort. However, a salesperson’s effort is naturally multidimensional. Hence, the compensation structure, in addition to affecting aggregate sales, may alter various dimensions of effort (Holmstrom & Milgrom, 1991). For example, Kishore et al. (2013) find that a switch from a discrete bonus to a commission plan increases effort on incentivized tasks but decreases effort on non-incentivized tasks. Thus, a change in compensation may affect not only aggregate sales but also other dimensions of performance.

⁵ Alternative views from that of Camerer et al. (1997) are presented in Farber (2005, 2008, and 2015). We thank an anonymous reviewer for bringing this to our attention.

In summary, it is unclear how a change in the quota frequency of the compensation scheme affects sales performance across multiple dimensions. Hence, through theoretical illustration and empirical validation via a field experiment, this study attempts to gain insights by examining salespeople’s behavior (amount of effort and product focus) regarding the compensation plan’s quota frequency.

3. Theoretical Model

The purpose of this section is to provide a parsimonious theoretical model that explains salespeople’s behavior in response to different quota frequency (quota cycles) in the compensation structure, while keeping total compensation (conditional on performance) constant. We begin by discussing the model setup and then illustrate how heterogeneous salespeople respond differently—in terms of both the quantity and type of effort—to a change in the quota frequency of the compensation scheme.

3.1. Model Setup

3.1.1. Agent’s Problem

A salesperson⁶ (agent) chooses to focus his or her selling effort on either a low-ticket product (L) or a high-ticket product (H) in periods $k = 1, 2, \dots$ (infinite-horizon).⁷ If the agent focuses on $\theta \in \{L, H\}$ in period k , a sales volume $\chi^\theta > 0$ is generated with probability $0 \leq p^\theta < 1$, and zero sales volume occurs with probability $1 - p^\theta$. The agent can also take the null action (\emptyset), where zero sales volume occurs with probability 1. The sales volume of each product H and L —by definition, high-ticket and low-ticket, respectively—is $\chi^H > \chi^L > 0$. Sales volumes across different periods are statistically independent, conditional on the agent’s choice of effort.

⁶ Hereafter, we use the term salesperson and agent interchangeably to refer to an employee hired by the firm to engage in its sales activities.

⁷ We define an agent’s choice to exert any effort (either on L or H versus no effort) as the quantity of effort; and conditional on exerting effort, the choice of the product focus (L or H) as the class of effort. Detailed explanations are provided in Section 3.2.1.

As is common in practice, the firm sets a number of periods ($N > 0$) to form an evaluation time window—typically referred to as a quota cycle. Hence, periods $1, \dots, N$ constitute the first quota cycle; periods $N + 1, \dots, N + N$ constitute the next quota cycle; etc. For example, if $k = \text{day}$ and $N = 30$, the quota cycle will be monthly. The agent receives compensation at the end of each quota cycle based on his or her performance. In addition, compensation depends only on sales within the specific quota cycle and not those in previous cycles; this is similar to sales compensation plans in practice. Formally, let $s_N(x_1, x_2, \dots, x_N) \geq 0$ denote an agent’s compensation conditional on the length of the quota cycle N and realized sales x_1, \dots, x_N within a quota cycle.

The agent’s total payoff is the sum of total compensation and the disutility (cost) of effort. Let $a_k \in \{a^H, a^L, a^\emptyset\}$ denote the agent’s action in the k -th period within the quota cycle of interest, where a^H and a^L indicate the agent’s actions focusing on H and L products, respectively, and a^\emptyset indicates the null action. The disutility from each action is given by $c(a^H) = c^H > 0$, $c(a^L) = c^L > 0$, and $c(a^\emptyset) = 0$, where $c^H \geq c^L > 0$. Hence, the agent’s expected total payoff in each quota cycle is

$$\mathbb{E} \left[s_N(x_1, x_2, \dots, x_N) - \sum_{k=1}^N c(a_k) \right]. \quad (1)$$

A policy of the agent specifies which action to take in each period, contingent on the agent’s actions and realized sales in previous periods within a quota cycle. Given that performance in previous quota cycles does not affect future compensation, the agent will repeat the same policy in every quota cycle. Specifically, in every quota cycle, the agent will employ a policy that maximizes **Equation (1)**. The agent’s optimal policy clearly depends on N , and let Σ_N denote the set of all agent-optimal policies, conditional on N .

3.1.2. Firm’s Problem

The firm chooses the structure of the compensation plan—specifically, the length of the quota cycle N —to maximize its per-period expected revenue. The agent, in response to the compensation plan, selects a policy that maximizes his or her payoff. Sales are realized, and the agent is compensated as described in the previous subsection. To specifically focus on our main research

question of interest—how the length of the quota cycle (N) affects agents’ behavior—we assume that other elements of the compensation scheme s_N are exogenously given to the firm and that the firm is allowed to choose only the length of quota cycle N .⁸

The firm chooses the quota cycle N that maximizes its per-period revenue μ_N , with the expectation that the agent will choose an agent-optimal policy from Σ_N .⁹ Formally, let

$$\mu_N = \sup \left\{ \mathbb{E}^\sigma [X/N] : \sigma \in \Sigma_N \right\}$$

denote the maximum expected per-period revenue as a function of N , where $X (= x_1 + \dots + x_N)$ is the agent’s total sales volume within the quota cycle, and $\mathbb{E}^\sigma [X/N]$ is the expected average sales volume under an agent-optimal policy σ .¹⁰

3.1.3 Commission-based Compensation Scheme

The main goal of this study is to examine how a change in the quota cycle N affects the incentive and, thus, the behavior of the agent. To separate out the effect of a change in N from other factors, we make the following assumptions.

Assumption 1. $s_N(x_1, \dots, x_N) = s_N(\tilde{x}_1, \dots, \tilde{x}_N)$ whenever $\sum_{k=1}^N x_k = \sum_{k=1}^N \tilde{x}_k$.

Assumption 1 requires that an agent’s compensation depends only on the total sales volume within a quota cycle. This assumption excludes, for example, a compensation scheme whereby an agent is rewarded whenever he or she sells a high-ticket (respectively, a low-ticket) product but receives zero

⁸ We abstract from optimality of the overall shape of the compensation structure (Holmstrom, 1979) to focus on the optimal length of the quota cycle, which was the main interest of the focal firm and, thus, this study’s field experiment setting. However, the shape of the focal firm’s compensation structure (shown in **Table 2** and **Figure 3**) closely resembles the right-continuous convex structure, as in the optimal contract derived by Holmstrom (1979).

⁹ For simplicity (i.e., stationarity), we assume that the firm maximizes per-period revenue, which is equivalent to the firm maximizing total revenue over multiple periods with a unit discount factor.

¹⁰ If there exists more than one agent-optimal policy, we assume that the agent chooses the one that maximizes the firm’s average sales volume—that is, the agent breaks a tie in favor of the firm whenever he or she is indifferent.

if the agent sells a low-ticket (respectively, high-ticket) product.¹¹ Assumption 1 (that compensation depends only on total sales volume), indeed, holds in many compensation systems observed in practice, including the focal firm for this study’s empirical analyses. Hereafter, for any quota cycle N , let $s_N(X)$ denote the agent’s compensation conditional on his or her total sales volume $X (= x_1 + \dots + x_N)$ within the quota cycle.

Assumption 2. $Ns_1(y) = s_N(Ny)$ for any N and $y \geq 0$.

The above assumption is necessary to compare the agent’s incentives under different quota cycles while keeping total compensation (conditional on performance) fixed. For example, suppose that the agent’s sales volume is identically y in all periods $k = 1, 2, \dots, N$. The agent’s total compensation for this series of performance is $Ns_1(y)$ if the quota cycle is one (the agent receives $s_1(y)$ in each period) but $s_N(Ny)$ if the quota cycle is N . The assumption simply indicates that if an agent’s performance is identically y in all periods across the two quota cycles, his or her total compensation is the same regardless of N . Without this assumption, the firm can potentially provide different compensation to the agent even when the agent’s performance is identical. Again, this study’s main objective is to examine the causal effect of quota frequency on salespeople’s behavior; thus, we need to compare the agent’s behavior with regard to a change in N while keeping everything else, including compensation (conditional on performance), constant.

The above two assumptions facilitate the theoretical analyses by imposing a restriction on the structure of s_N . Under Assumptions 1 and 2, for any quota cycle N and total revenue $X = x_1 + \dots + x_N > 0$, an agent’s compensation can be represented as

$$s_N(X) = Ns_1(X / N) = \frac{Ns_1(\bar{x})}{X} X = \frac{s_1(\bar{x})}{\bar{x}} X, \quad (2)$$

¹¹ By offering the agent this kind of compensation scheme, the firm effectively assigns (or promotes) him or her to a position such that the agent *must* focus solely on high-ticket products (respectively, low-ticket products). Any realization of $x_k = \chi^L$ (respectively, $x_k = \chi^H$) will indicate that the agent was disobedient in period k , resulting in zero compensation as a punishment. Assumption 1 allows us to abstract from the problem of assigning and/or promoting agents across different positions.

where $\bar{x} = X / N$ is the average sales volume (per period) within the quota cycle. Notice that the agent’s compensation in **Equation (2)** can be expressed as proportional to total sales X —that is, a commission-based compensation plan with a commission rate $s_1(\bar{x})/\bar{x}$, which is a function of the average sales volume within a quota cycle. Hereafter, we denote the commission rate as

$$\beta(\bar{x}) = \frac{s_1(\bar{x})}{\bar{x}},$$

such that the agent’s compensation is

$$s_N(X) = \beta(\bar{x})X \tag{3}$$

for any quota cycle N . By structuring and rearranging the compensation plan as in **Equation (3)**, we can examine changes in the agent’s behavior with regard to changes in the quota cycle N , while keeping other elements of the compensation plan constant.

3.2. Analysis

There are four parameters, (p^H, p^L, c^H, c^L) , that summarize the agent’s ability. We focus on the comparative static analyses with respect to p^H , the parameter representing an agent’s ability to promote high-ticket products¹²—that is, how the firm’s optimal quota cycle differs for agents with different levels of p^H . High-ticket products typically are sophisticated and technologically advanced, needing substantial information for usage and handling. Salespeople help provide such information to potential consumers. Moreover, many new high-ticket products are experience goods, and consumers often rely on salespeople’s assessment of their value before purchase. Hence, a consumer requires more sales assistance when deciding whether to purchase high-ticket products than when considering low-ticket products. The quality of such sales assistance is highly dependent on a

¹² Equivalently, we can keep p^H constant and examine comparative statics with respect to disutility c^H , which does not change the qualitative feature of the model’s predictions. More generally, we can conduct comparative statics with respect to any of the four parameters (p^H, p^L, c^H, c^L) , holding the other three constant. The qualitative feature of the model hinges only on what is more productive (or what incurs less disutility) between high- and low-ticket products across different type of agents.

salesperson’s ability, which is typically heterogeneous across salespeople. In contrast, low-ticket products do not require as much sales assistance, and the ability to sell them is relatively homogeneous across salespeople.

In what follows, we refer to a salesperson with a high value of p^H as a ‘high-type’ salesperson and to salesperson with a low value of p^H as a ‘low-type’ salesperson, and we analyze the effect of changing N on high-type and low-type salespeople separately. Note that the notations L and H denote low-ticket and high-ticket products, respectively, while low-type and high-type agents represent those with low and high values of the parameter p^H (the ability to promote and sell high-ticket products), respectively.

3.2.1. Agent’s Choice of Effort and Product Focus

The agent’s decision consists of two aspects of effort: 1) the quantity and 2) the class of effort. First, the agent chooses whether to exert effort, defined as the quantity of effort; hence, the quantity of effort in each period is binary (zero or one). Second, conditional on exerting effort, the agent also chooses which product type to focus on (either high-ticket or low-ticket), defined as the class of effort.

The firm cares about both the quantity and the class of effort. Note that the firm may not want the agent to focus solely on high-ticket products. First, for an agent whose $p^H \chi^H < p^L \chi^L$, the firm prefers him or her to exert effort on low-ticket products because the expected revenue is higher for low-ticket than for high-ticket products. Second, even for an agent whose $p^H \chi^H > p^L \chi^L$, the firm may not want him or her to divert effort to high-ticket products if such a change is accompanied by a significant reduction in the quantity of the agent’s effort. Simply put, the firm, in each period, wants agents to exert effort versus no effort. In addition, the firm wants agents to exert effort consistent with their types. That is, the firm wants agents who are effective at selling high-ticket products (salespeople with high p^H) to focus on those products and agents who are effective at selling low-ticket products (salespeople with low p^H) to focus on those.

3.2.2. Firm’s Optimal Quota Frequency for Low-Type Agents

This section examines the provision of incentive to a low-type sales agent (a salesperson with low p^H). The following proposition shows that it is optimal for the firm to set the most-frequent quota cycle ($N = 1$) for a low-type agent.

Proposition 1. *Suppose Assumptions 1 and 2, and suppose that $\beta(\cdot)$ is non-decreasing, piecewise-continuous, and non-negative.*

- (i) *There is $\underline{p}^H > 0$ such that, whenever $p^H \in [0, \underline{p}^H]$, $\mu_H \geq \mu_N$ for any integer $N > 1$.*
- (ii) *Suppose that the firm chooses $N > 1$ and that $\beta(\chi^L/N) < c^L/(p^L\chi^L) < \beta(\chi^L)$. There is $\underline{p}^H > 0$ such that $\mu_H > \mu_N$ whenever $p^H \in [0, \underline{p}^H]$.¹³*

The formal proof of the above proposition is in Appendix A. There are two points to note. The proposition imposes no restriction on the compensation scheme (i.e., the commission rate β) beyond monotonicity and piecewise continuity, which are observed in virtually all commission schemes used in practice. In addition, the proposition assumes that agents are fully rational and forward-looking. This stands in contrast to other existing theories that explain the effectiveness of frequent goals (quota cycles) based on psychological/behavioral assumptions (Hull, 1932; Hull, 1938; Heath, Larrick, & Wu, 1999; Kivetz, Urminsky, & Zheng, 2006; Nunes & Drèze, 2006; Cheema & Bagchi, 2011; Gal & McShane, 2012; Zhang & Gao, 2016).

The intuition for this proposition is relatively straightforward. When p^H is close to zero, under any quota cycle, it is always suboptimal for the agent to exert effort on high-ticket products. Hence, the firm is concerned only with maximizing the total quantity of the agent's effort, without worrying about the class of effort. In other words, it is in the firm's best interest to make the agent exert effort on low-ticket products (instead of the null action) as much as possible (Observations 1 and 2 in the proof of Proposition 1-i; see Appendix A for details).

¹³ Proposition 1-i shows that $N=1$ generates at least weakly more sales volume for a low-type agent than does any other N . However, Proposition 1-i alone does not rule out the possibility that μ_N is independent of N , and, thus, $N=1$ is only *vacuously* optimal. Proposition 1-ii identifies the condition under which $N=1$ generates *strictly* more sales volume.

Relatedly, in terms of intertemporal dynamics, if $N > 1$, it occurs with positive probability that the agent's effort fails to realize in the early periods of the quota cycle. In this case, the agent loses hope about meeting quota and obtaining high compensation at the end of the quota cycle. Consequently, the agent gives up and stops working in later periods. In contrast, if $N = 1$, bad luck in early periods does not affect the agent's incentive in later periods, and, thus, the agent exerts effort in all periods. Hence, the shortest quota cycle ($N = 1$) is the most effective plan for incentivizing low-type sales agents (see proof of Proposition 1-ii in Appendix A for details).

Let us illustrate an agent's optimal policy when the quota cycle is $N = 1$. Suppose that the agent exerts effort on high-ticket products. The agent incurs disutility c^H , regardless of the sales outcome, which will realize to χ^H with probability p^H or to zero with probability $1 - p^H$. Hence, the agent's expected utility from exerting effort on high-ticket products is

$$u^H := p^H \beta(\chi^H) \chi^H - c^H.$$

Similarly, the expected utility from exerting effort on low-ticket products is

$$u^L := p^L \beta(\chi^L) \chi^L - c^L.$$

The agent chooses to exert effort on high-ticket products only if $u^H \geq u^L$ so that

$$p^H \geq A(p^L) \equiv \frac{\beta(\chi^L) \chi^L}{\beta(\chi^H) \chi^H} p^L + \frac{c^H - c^L}{\beta(\chi^H) \chi^H}.$$

On the other hand, the agent chooses to exert effort on low-ticket products only if $p^H \leq A(p^L)$.

Figure 2a describes the optimal action that maximizes the agent's payoff for each combination of (p^L, p^H) . The agent focuses on high-ticket products if his or her (p^L, p^H) lies in the horizontally shaded area and on low-ticket products if his or her (p^L, p^H) lies in the gridded area. The agent will take the null action in all other cases because both u^H and u^L are negative. Simply put, the agent exerts effort on the product types that are consistent with his or her relative ability (p^H to p^L ratio) if the minimum expected utility of exerting effort is satisfied.

Figure 2b is identical to **Figure 2a**, except for the line initiated from the origin. This line is the collection of points (p^L, p^H) at which both high- and low-ticket products generate the same

expected revenue for the firm. If an agent's (p^L, p^H) lies in the area above (below) the line, the firm prefers the agent to focus on high-ticket (low-ticket) products. The gridded area in **Figure 2b** represents the case in which the incentives of the agent and the firm are aligned under the quota cycle $N = 1$. The agent focuses on low-ticket products to maximize his or her expected utility, which also maximizes the firm's expected revenue.¹⁴

3.2.3. Firm's Optimal Quota Frequency for High-Type Agents

Proposition 1 shows that setting the most-frequent quota cycle ($N = 1$) is effective in motivating low-type agents, maximizing the quantity of their effort and, thus, revenue. However, setting a frequent quota cycle may not be effective in incentivizing high-type agents. See the diagonally shaded triangular area in **Figure 2b**. It is optimal for the firm that agents in this area focus on high-ticket products; yet they focus on low-ticket products under the quota cycle $N = 1$. Hence, the firm may use a longer quota-cycle compensation scheme to divert agents' efforts from low-ticket to high-ticket products.

To find the optimal quota cycle for high-type agents, ideally, one may attempt to solve for the optimal N for each p^H and then examine how the optimal N changes in response to a change in p^H . Unfortunately, this problem is generally intractable, and, thus, it is difficult to obtain a closed-form solution. The difficulty stems from the richness of the agent's feasible policies, whose number increases exponentially in N . Furthermore, the agent's dynamic optimization problem lacks stationarity because of the clear *deadline effect*: the agent's problem in periods near the end of a quota cycle is substantially different from that in earlier periods. These technical issues prevent an analytical prediction of how the agent responds to a change in N , which is necessary to characterize a firm-optimal quota cycle.

To circumvent these technical issues, we make two compromises. First, instead of a general model, we restrict attention to an example (a single-tier quota-commission compensation scheme) that makes the problem more tractable. Second, rather than seeking to fully characterize the firm's

¹⁴ The gridded area includes the segment $\{(p^L, p^H): 0 < p^L < c^L / \beta(\chi^L)\chi^L \text{ and } p^H = 0\}$ on the horizontal axis. These agent types take the null action under any quota frequency, and, thus, the quota frequency of $N=1$ is vacuously optimal for the firm.

optimal quota cycle, we aim to prove that the optimal quota cycle, under certain conditions, is never equal to one ($N \neq 1$ and, thus, $N > 1$) for high-type agents.

Hereafter, the focus of the compensation structure is on a single-tier quota-commission scheme¹⁵:

$$\beta(\bar{x}) = \begin{cases} b & \text{if } \bar{x} \geq q \\ 0 & \text{if } \bar{x} < q \end{cases}$$

for commission rate $b > 0$ and quota $q > 0$. The agent obtains commission pay (commission rate $b > 0$) only if he or she achieves the quota $\bar{x} \geq q$; otherwise, the agent earns only the base salary and no commission pay.

To avoid trivial and/or uninteresting cases, we also impose the following two assumptions (See Section 3.2.2 for the definition of u^L and u^H).

Assumption 3-A. $p^H \chi^H > q > p^L \chi^L$.

Assumption 3-B. $u^L > u^H > 0$.

Assumption 3-A is, in fact, the combination of two conditions: (i) $p^H \chi^H$ is larger than $p^L \chi^L$; and (ii) the quota q lies in between. The first condition implies that the firm wants the agent (with characteristics p^H and p^L) to focus on high-ticket products. The second condition indicates that the agent likely achieves quota q (on average) if he or she focuses on high-ticket products in all periods, whereas the agent who focuses only on low-ticket products is unlikely to achieve quota (again, on average).

Assumption 3-B indicates that, contrary to the firm's wishes, the agent chooses $a = a^L$ when $N = 1$. From Assumption 3-A, the agent, by focusing on a high-ticket product, attains greater sales, on average. However, he or she incurs greater disutility, which negatively offsets higher sales, and, thus, the agent chooses to focus on low-ticket products. Note that Assumption 3-B implicitly requires that $\chi^H > \chi^L \geq q$ (otherwise, it is impossible for both u^H and u^L to be positive).

Assumptions 3-A and 3-B, collectively, imply that the firm wants the agent to exert effort on high-ticket products, but the agent focuses on low-ticket products when $N = 1$, even though high-

¹⁵ We use this incentive structure for illustration because it is consistent with the theoretical incentive structure in **Equation (3)** and resembles the incentive plan utilized by the focal firm for our empirical analysis.

ticket products return higher expected revenue. Then, the natural question is whether the firm, by setting a longer quota cycle ($N > 1$), can divert the agent’s effort to high-ticket products and increase its expected revenue.

Proposition 2. *Suppose that the compensation structure takes the form of a single-tier quota-commission scheme. Under Assumptions 1, 2, 3-A and 3-B, a quota cycle of $N = 1$ is never optimal for the firm.*

The formal proof in Appendix B shows that $\mu_N > \mu_1$ for a sufficiently long quota cycle N . The basic intuition for the proposition is as follows. Suppose that the firm sets a sufficiently long quota cycle. If the agent always focuses on low-ticket products (as is the case under the quota cycle $N=1$), he or she almost surely ends up with an average sales volume approximately equal to $p^L \chi^L < q$ (by the law of large numbers) and, thus, zero compensation. This clearly is suboptimal for the agent.

In fact, when N is sufficiently large, the (high-type) agent should divert some effort toward high-ticket products to achieve the quota. The law of large numbers eliminates all the randomness in the agent’s performance, and, thus, the agent will have no chance of achieving quota unless he or she exerts some effort on high-ticket products. The diversion of effort would also increase the firm’s revenue because $p^H \chi^H > p^L \chi^L$ by assumption. See Appendix B for details.

3.2.4. Theoretical Prediction

Based on the results of the theoretical analyses provided in this section, **Table 1** presents the overall prediction of product focus and revenue by agent types when a firm changes from a long quota-cycle (e.g., $N=30$) to a short quota-cycle (e.g., $N=1$) compensation structure. As shown in Proposition 1, the low-type agents improve on revenue by increasing their quantity of effort and, thus, the number of products sold. However, their product focus (on low-ticket products) remains unchanged. In contrast, the high-type agents, as shown in Proposition 2, change their focus from high- to low-ticket products, decreasing total revenue. The subsequent sections show the detailed procedures and results of a field experiment that validates the theoretical predictions.

4. Institutional Details and the Field Experiment Design

To validate the theoretical claims in the previous section, we conducted a field experiment with full-time sales employees, varying a firm’s compensation quota frequency. The focal firm is a highly regarded retail chain operating 94 stores in Sweden. It sells mostly accessories for cellular phones and home electronics (e.g., networking accessories, headsets and phone cases) and parts for consumer electronics and home appliances (e.g., semiconductors and switches). It also sells small-to-medium-sized consumer electronic goods, such as data-storage devices, network appliances, DVD players, and wireless routers. Product prices range from less than \$1 to \$500 or more, with an average price of approximately \$20. All of the stores are company-owned, and the firm employs a direct sales force of about 350 salespeople at any given time across its stores. The compensation plan for sales employees consists of a fixed salary plus a variable commission on sales. The commission rate (and, thus, the commission amount) is determined by sales performance, measured in average sales per hour (SPH).

Table 2 shows the details of the variable component of the compensation plan, with five levels of commission available to salespeople. For example, if a salesperson’s average SPH were \$150 at the end of the evaluation period, he or she would receive a commission of 0.27% for every dollar of sales. If a salesperson’s average SPH were \$250 or more, he or she would receive the highest commission level of 2.0%. Note that the quotas are in average SPH instead of in absolute amounts. Because of this characteristic, along with the discrete nature of the tiered commission levels, a salesperson’s variable pay would have a kink at each tier (quota) level, which resembles a combined quota-commission and quota-bonus scheme (**Figure 1d**). **Figure 3a** illustrates the level of variable pay for a salesperson assigned 140 hours a month. The figure shows that, as a salesperson achieves each quota level, he or she receives a step jump in pay due to the discretely accelerating commission rates. A salesperson would make \$1,000 in variable pay if his or her monthly sales totaled \$50,000.

The field experiment and, thus, the change in the compensation plan took effect on May 1, 2015. Holding everything else constant—including the commission rate per quota achieved and the quotas in terms of average SPH (**Table 2**)—only the evaluation period changed (from monthly to daily) on May 1. That is, up until the end of April 2015, the firm assessed each salesperson’s commission

rate by summing all the sales that the individual made in a month and dividing them by all the hours that he or she worked within that month. Then, starting on May 1, the firm began to evaluate the commission rate daily. **Figure 3b** illustrates commission pay as a function of daily sales for a salesperson assigned seven hours a day. The overall shapes of **Figure 3a** and **Figure 3b** are similar, as only the frequency of evaluation changed, while the commission rate and the quotas in **Table 2** remained the same.

In addition to changing the compensation structure for employees across the firm's stores (the treatment group), we also arranged for several stores and their salespeople to experience no changes in the compensation plan during the experiment (the control group). Hence, the difference in sales performance between the treatment and the control groups identifies the magnitude of the treatment effect (the daily-quota plan), taking into account any normal/temporal changes (e.g., seasonality or firm-level advertising) in sales that would have occurred regardless of the change in the compensation plan.

With help from management, we chose stores for the control group that embodied a representative sample of stores across the geographical areas of the country.¹⁶ The majority of Sweden's population resides in the southern tip of the country, concentrated in the suburbs and city centers of the three most populous cities, Stockholm, Gothenburg, and Malmö. There were two main challenges in choosing the control stores.

First, to avoid complications in implementing the changed compensation plan, the focal firm initially did not want any control stores. In addition, the firm's management was extremely concerned about fairness across employees. Many members of management (including the sales director, the information technology director, and the vice president of operations) had risen through the ranks, starting out as in-store salespeople, and fairness was one the firm's primary human

¹⁶ The firm had only eight stores in the entire central and northern parts of Sweden; we omitted these stores from our analysis, leaving us with 86 stores.

resources (HR) policies.¹⁷ Thus, they deemed the concept of having some employees on a different compensation plan to be extremely inappropriate.

Second, control stores needed to be comparable (with similar characteristics) to surrounding stores, but not too close in geographical proximity to treatment stores in order to avoid the “water cooler effect.”¹⁸ That is, we did not want the salespeople in the control and treatment groups to communicate with each other, as such communication could have biased the outcome of the field experiment. For the control group, we ended up randomly selecting five stores (consisting of 26 salespeople) from a set of stores not in very close proximity to other stores, but still in the metropolitan areas of Stockholm, Gothenburg, and Malmö. To further avoid the water cooler effect, we made sure that there were no major sales training programs or conferences around the time of the field experiment, as these events could have led to a spillover of information. In addition, we made sure that there were no employee transfers between the treatment and control stores during the experiment. The focal firm’s management was also very concerned about the water cooler effect, but for a different reason. As mentioned above, the firm prided itself on its HR policy of one-for-all. This is the main reason that, for the control group, we were limited to only five stores for a period of one month.

For the field experiment to be valid, it is important for the performance of the control and treatment groups to be similar, especially with regard to the variability in performance over time. **Table 3** shows the summary statistics of both groups in April 2015, the pre-intervention period. Both the mean and the standard deviation are quite similar across the two groups. **Table 3** shows only that sales performance of the control group and that of the treatment group are from similar distributions. However, and more importantly, for the empirical analysis to be valid, the temporal

¹⁷ Sweden is often cited as having one of the highest effective tax rates in the world. It is also known for having generous social security benefits, including child care, health care, housing allowances, and welfare. Sweden also ranks high with regard to gender equality (World Economic Forum, 2014). Hence, fairness is a relatively strong social norm in Sweden.

¹⁸ The term “water cooler effect” refers to the phenomenon in which employees gather around the office water cooler to talk. In this study’s context, “water cooler effect” refers to the flow of information that potentially disrupts the motivation of salespeople and, thus, the direction and the magnitude of the effect of the experiment treatments.

trend in the two groups must match. For verification, we perform several placebo tests in Section 6.

Table 4 shows the average SPH for the control and treatment groups across April and May 2015. Once again, the compensation plan changed from a monthly- to a daily-quota plan as of May 1. The benefit and importance of having a control group is clear. Given the 10% improvement in sales performance (i.e., an increase in average SPH from \$149.06 to \$163.96) across the two periods for the treatment group, one can conclude that the plan change was immensely successful and that the daily-quota plan outperformed the monthly-quota plan. However, taking into account the differences in performance of the control group across the two periods, that conclusion is not so obvious. There seems to be only a marginal effect—a 10.0% increase for the treatment group compared with a 9.1% increase (i.e., an increase in average SPH from \$149.17 to \$162.75) for the control group. The above analysis shows only an aggregate result of the field experiment, not taking into account individual heterogeneity (both in terms of permanent heterogeneity and responsiveness to incentives) and daily aggregate shifts in demand. The next section provides an empirical model that takes these features into account and considers sales as a function of a salesperson’s effort, conditional on the compensation structure.

5. Empirical Model

Sales performance Y_{id} of salesperson i on day d is a multiplicative function of the salesperson-specific effects α_i , common daily time trends γ_d , the compensation structure z_{id} , and an idiosyncratic shock ε_{id} such that

$$Y_{id} = \exp\left(\alpha_i + \gamma_d + \delta z_{id} + \varepsilon_{id}\right), \quad (4)$$

where z_{id} is a binary variable with a value of one if salesperson i is in the treatment group and day d is in the treatment period. The parameter α_i represents unobserved individual heterogeneity that is constant over time, and the parameter γ_d represents any intertemporal variations that are common across all salespeople on a particular day. Examples include seasonal fluctuations in demand (e.g.,

due to the weather) or the firm’s other marketing activities (e.g., advertising) that affect all salespeople equally on day d . The parameter δ represents any increase (or decrease) in salesperson i ’s effort as a result of the change in the compensation structure (i.e., quota frequency). The idiosyncratic shock ε_{id} represents any other elements that affect sales, such as luck (either good or bad). The idiosyncratic shocks are assumed to be heteroskedastic and independently, identically, and normally distributed within salespeople over time with mean zero and variance σ_i^2 . The logarithmic transformation of **Equation (4)** leads to the empirical model:

$$\log(Y_{id}) = y_{id} = \alpha_i + \gamma_d + \delta z_{id} + \varepsilon_{id}. \quad (5)$$

The identification of the treatment effect results from any difference in performance between the treatment and the control groups, controlling for any natural trends common to both groups.¹⁹ Technically, identification of the treatment effect can occur just by cross-sectional analysis, using data only from periods after the treatment (May 1), assuming homogeneity—that is, the sales of the treatment and control groups were identical before the treatment. Our sample size of 337 employees, although quite large for a field study, is not sufficiently large for random assignments in treatment conditions to eliminate individual fixed effects. The empirical approach in **Equation (5)** allows the use of full information from the data to better control for individual heterogeneity, providing robust estimates of the treatment effect.

6. Results

First, we discuss the effect of quota frequency on overall sales performance. Then, we perform robustness tests to validate our empirical results. Finally, we check for the effects of quota frequency on other dimensions of performance: the quality of effort (product returns) and the class of effort (product focus).

6.1. Quota Frequency and Sales Performance

¹⁹ The difference-in-differences method (Card & Krueger, 1994) is frequently used to mimic an experimental research design with naturally occurring data.

The first column of **Table 5** shows the result of **Equation (5)** with the logarithm of SPH per day as the dependent variable. Consistent with the model-free results in **Table 4**, in aggregate, there is no significant increase in sales performance under the daily-quota scheme compared to that of the monthly-quota scheme. The second column of **Table 5** shows the results of a deviant model of **Equation (5)**. Specifically, to examine heterogeneity in responsiveness across agents, we allow for different slope parameters by segments of salespeople such that

$$\log(Y_{id}) = y_{id} = \alpha_i + \gamma_d + \sum_{r=1}^R \delta_r I_{(i \in S_r)} z_{id} + \varepsilon_{id}, \quad (6)$$

where $I_{(i \in S_r)}$ is an indicator function that equals one if salesperson i is a member of Segment r , S_r , and δ_r is the corresponding segment-level parameter. The segmentation is via a quartile split with regard to sales performance before the treatment period.²⁰ An interesting pattern (with regard to responsiveness by salespeople of different performance levels) emerges. While the frequent (daily) quota plan has a positive effect on low-performing salespeople, it has a negative effect on high-performing salespeople. In terms of magnitude, the lowest-performance segment has a positive and significant effect, with an 11.7% increase in sales performance.²¹ In contrast, the highest-performance segment has a negative and significant effect, with an 8.1% decrease in sales. Although directionally consistent with the above assessment, the two mid-quartile segments do not show any statistically significant effects.

Frequent quotas (shorter-term goals) seem to have a positive effect on low-performing salespeople. By definition, low-performing salespeople have greater disutility of effort or are less efficient for a given amount of effort, or both.²² Under a monthly-quota plan, a salesperson who had

²⁰ Because we use the difference in sales performance from the control group to identify the treatment effect, we control for the regression-to-the-mean phenomenon in the data. This pattern in the data, if any, would be present for both the control and treatment groups. We thank an anonymous reviewer for this comment.

²¹ Because of the logarithmic transformation of our dependent variable, the magnitude of the estimated treatment effect is equal to an 11.7% increase in sales performance, using the transformation formula $\exp(0.111) - 1 = 0.1174$.

²² In Section 3, we defined low-performing (or low-type) salespeople as those who are less effective at selling high-ticket products.

bad luck (reduced sales) in the earlier part of the month will give up in the later days of the month because there is virtually no chance of meeting quota by the end of the month. This would not be the case under a daily-quota plan, as there is a fresh start every day, in which past performance does not affect current payoff and, thus, does not distort current motivation.²³ This is consistent with the results of the theoretical model in Section 3.

The education literature has found that frequent testing results in better outcomes (see Bangert-Drowns, Kulik, & Kulik (1991) for an extensive survey). Furthermore, studies in the psychology literature have found that breaking up a main goal into multiple subgoals results in more-favorable outcomes (Heath, Larrick, & Wu, 1999). Similarly, the marketing literature has shown that frequent goals result in favorable outcomes (Gal & McShane, 2012; Zhang & Gao, 2016). In the sales management literature, Chung, Steenburgh, & Sudhir (2014) have explored the concept of quota frequency and found that quarterly bonuses help salespeople to achieve the annual quota. However, their analysis was based on counterfactual simulations, using estimates from their structural model, and not inferred directly from the data.²⁴ To the best of our knowledge, sales-quota frequency has never been directly analyzed from empirical data, let alone through a field experiment. Our results show that frequent quotas benefit low-performing salespeople, much like frequent classroom testing helps improve the performance of low-ability students.

In contrast to the findings with regard to low-performing salespeople, frequent quotas lead to a decrease in sales for high-performing salespeople, consistent with the theoretical prediction given in Section 3 (see **Table 1**). To the best of our knowledge, this is a novel finding that has not been previously documented. We discuss this result, in detail, in Section 6.4.

²³ In addition to negating salespeople’s giving-up behavior after a series of unsuccessful days, the daily quota plan can also circumvent the starting problem (Heath, Larrick, & Wu, 1999).

²⁴ After estimating a structural model of heterogeneous sales force behavior responding to annual and quarterly quotas, Chung, Steenburgh, and Sudhir (2014) demonstrate, using counterfactual simulations, that removing quarterly quotas (and keeping only annual quotas) leads to a greater decrease in the performance of low-performing salespeople. In contrast, in this research, we directly test the effect of change in quota frequency on sales performance and provide direct empirical support for this conjecture. In addition, we investigate the effect of quota frequency not only on aggregate sales, but also on various other dimensions of performance, such as the quality and type (class) of effort, by monitoring and measuring product returns and changes in the product focus.

6.2. Robustness Tests

As mentioned in Section 4, for the empirical analysis to be valid, the time trend must be parallel between the control and treatment groups. Because the control group was smaller than the treatment group and was chosen quasi-randomly (randomly chosen from a set of stores), we need to verify whether the parallel trend assumption is valid. For verification, we perform several placebo tests²⁵ by pretending that a treatment occurred for the treatment group during the non-treatment period (April). The details of the placebo tests are as follows. First, using only April data, we choose each week (1st, ... , 4th) and imagine that a treatment had been implemented for the treatment group. Then, we perform the same analysis using **Equation (6)**. The results are shown in **Table 6**. Regardless of the week used for the false treatment period, all of the parameter estimates are very small and statistically insignificant. These results indicate that the treatment effect (in **Table 5**) is a result of an *actual* treatment and not that of any other systematic temporal changes, different from the treatment to the control groups.

6.3. Quota Frequency and Product Returns

The focal firm prides itself on being known for its excellent customer service. It trains its salespeople to know the technical specifications of its entire range of products, as well as their applications. In addition, the firm ensures that salespeople regularly undergo a significant amount of customer-service training, well above the industry norm in Sweden. Hence, management was concerned that the daily-quota plan, while potentially increasing short-term motivation, might be harmful if it resulted in salespeople aggressively selling unnecessary products to customers, which could result in an increase in returned merchandise. This concern was further aggravated by the fact that the firm did not penalize salespeople for product returns by reducing their compensation in either the daily- or the monthly-quota plans.

To examine whether there was a change in returns, we tracked all returned products and mapped them back to their original sales to create the variable, returns-to-sales (RTS) ratio. For example, if a salesperson sold \$1,000 worth of goods on April 1, \$30 worth of which were eventually returned,

²⁵ We thank the Department Editor for suggesting this robustness check.

the RTS ratio would be 0.03. The RTS ratio, which is normalized by total sales, provides insights into problems that may be associated with service quality. **Table 7** shows the results of **Equation (5)** and **Equation (6)**, using the logarithm of the RTS ratio per day as the dependent variable. No meaningful effect seems to exist as, in both the homogeneous and heterogeneous models, none of the parameters is statistically significant. This suggests that an aggressive compensation plan alone does not necessarily lead to a decrease in the quality of effort or to unethical behavior. It is likely a combination of an aggressive compensation plan and other factors, such as a misguided company culture, that drives such behavior (Zoltners, Lorimer, & Sinha, 2016).

6.4. Quota Frequency and Product Focus

The results in Section 6.1 show that a frequent (daily) quota plan can have both a positive and a negative effect on sales performance, depending on the types of salespeople. To better understand the behavioral change among heterogeneous salespeople, we conduct the following analyses. First, we run **Equation (6)** using the logarithm of the number of products sold per hour as a dependent variable. The first column of **Table 8** shows the results. The number of products sold increased for the lowest-performing segment under the daily-quota plan. This indicates that the main reason for low-performing salespeople's increase in sales revenue under a frequent quota plan is that the sheer raw number of products sold increased, which is consistent with the theoretical prediction (an increase in the quantity of effort) in **Table 1**.

How about the types of products sold? The second column of **Table 8** presents the results of **Equation (6)** with the logarithm of the average price of products sold per day as the dependent variable. The average price of products sold decreased for the highest-performing segment under the daily-quota plan. This indicates that high-performing salespeople under a frequent quota plan tend to focus on the sale of low-ticket products. This result is also consistent with the theoretical prediction (a change in the class of effort and, thus, the product focus) in **Table 1**. The fact that high-performing salespeople are selling low-ticket products is not attractive, as the firm wants these salespeople to focus on high-ticket, high-value-added products that generate more profits. To the

best of our knowledge, there has been no research, to date, that examines the causal relation between frequency of goals and an agent’s type (product focus) and quality (product returns) of effort.

Overall, this study’s results provide sound empirical evidence—supported by theory—of various effects of quota frequency that give substantive insights into the use of quotas. While having frequent quotas in a compensation system may increase absolute sales, especially for low-performing salespeople, frequent quotas tend to induce high-performing salespeople to focus on low-ticket, low-margin items. This may result in a decrease in sales of high-value-added products, thus hurting the firm’s profits.

7. Conclusion

Monetary incentives, in the form of conditional payments based on performance, are one of the key instruments that organizations use to motivate their employees. These incentives are especially important in the domain of personal selling. A sales force compensation system typically consists of a fixed salary plus a variable payment conditional on the salesperson achieving a certain threshold of performance—a sales goal—referred to as a sales quota. Despite the common use of quotas, we do not fully understand their role, especially with regard to their temporal frequency. Hence, this study examines the causal effect of a compensation system’s quota frequency on various dimensions of performance for different types of salespeople.

To explore the role of quota frequency, we first develop a theoretical model of salespeople’s behavior in response to different lengths in a quota cycle. Our model takes into account many realistic elements, such as salespeople’s multi-dimensional effort, heterogeneity in ability, product focus, and forward-looking behavior. To empirically validate our theoretical predictions, we collaborated with a major Swedish retail chain to conduct a field experiment, varying the compensation structure of full-time salespeople. Specifically, holding everything else constant, we changed the sales force compensation scheme from a monthly- to a daily-quota plan. Because the quota was in the form of average sales per hour, the only change was the increase in quota frequency.

The results of the field experiment show that an increase in quota frequency improves sales performance mainly for low-performing salespeople, by preventing them from giving up when

confronted with early negative sales shocks within a quota cycle. Under a daily-quota plan, every day is a fresh start; thus, salespeople’s motivation is intact throughout the month, whereas under a monthly plan, salespeople will give up later in the month if they are too far away from (and, thus, have no chance of meeting) the quota. In contrast, there are negative effects of a frequent quota plan on high-performing salespeople, as they alter their product focus from high-ticket to low-ticket products, resulting in a decrease in sales of high-value-added products and, thus, decreasing the firm’s profits.

In summary, this study uses two methodologies to provide a comprehensive look into the role of quota frequency on various dimensions of sales force performance. The study’s findings will be valuable for organizations as they design their sales compensation systems. While reducing the time horizon of the quota may better motivate low-performing salespeople, organizations need to be mindful of the unintended consequences of such a move for the high performers. More importantly, organizations should also understand the overall impact of changing the time period for evaluating salespeople’s performance on their quality and type of effort.

There are some limitations to note. Because of concerns over fairness—that is, managers of the field experiment’s focal company were concerned about treating some employees differently from others—the control group was deployed for only five weeks. Naturally, a study that could maintain a control group for longer periods and restrict the flow of information from the treatment group would enable researchers to better analyze the long-term effects of quota frequency. Furthermore, this study was a one-time intervention and, thus, could not examine sequence or order effects. Finally, the venue was in Sweden, a country well known for its high tax rate and generous social security programs. Accordingly, fairness and a sense of community in that society is a prominent social norm. One could speculate that this study’s findings—both their direction and magnitude—would be more concrete in societies in which individualism is more the social norm. Although not addressed in this study, these topics would be exciting areas for future research.

References

- Amir, O., & Ariely, D. (2008). Resting on laurels: The effects of discrete progress markers as subgoals on task performance and preferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1158–71.
- Bangert-Drowns, R. L., Kulik, J. A., & Kulik C-L. C. (1991). Effects of frequent classroom testing. *Journal of Educational Research*, *85*(2), 89–99.
- Camerer, C., Babcock, L., Loewenstein, G., & Thaler, R. (1997). Labor supply of New York City cabdrivers: One day at a time. *Quarterly Journal of Economics*, *112*(2), 407–441.
- Card, D., & Krueger, A. B. (1994). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review*, *84*(4), 772–793.
- Cheema, A., & Bagchi, R. (2011). The Effect of goal visualization on goal pursuit: Implications for consumers and managers. *Journal of Marketing*, *75*(2), 109–123.
- Chung, D. J., Steenburgh, T., & Sudhir, K. (2014). Do bonuses enhance sales productivity? A dynamic structural analysis of bonus-based compensation plans. *Marketing Science*, *33*(2), 165–187.
- Darmon, R. (1997). Selecting appropriate sales quota plan structures and quota-setting procedures. *Journal of Personal Selling Sales Management*, *17*(1), 1–16.
- Farber, H. S. (2005). Is tomorrow another day? The labor supply of New York City cabdrivers. *Journal of Political Economy*, *113*(1), 46–82.
- Farber, H. S. (2008). Reference-dependent preferences and labor supply: The case of New York City taxi drivers. *American Economic Review*, *98*(3), 1069–1082.
- Farber, H. S. (2015). Why you can't find a taxi in the rain and other labor supply lessons from cab drivers. *Quarterly Journal of Economics*, *130*(4), 1975–2026.
- Fishbach, A., & Dhar, R. (2005). Goals as excuses or guides: The liberating effect of perceived goal progress on choice. *Journal of Consumer Research*, *32*(3), 370–77.
- Fishbach, A., Dhar, R., & Zhang, Y. (2006). Subgoals as substitutes or complements: The role of goal accessibility. *Journal of Personality and Social Psychology*, *91*(2), 232–42.
- Gal, D., & McShane, B. B. (2012). Can small victories help win the war? Evidence from consumer debt management. *Journal of Marketing Research*, *49*(4), 487–501.

- Heath C., Larrick, R. P., & Wu, G. (1999). Goals as reference points. *Cognitive Psychology*, 38(1), 79–109.
- Holmstrom, B. (1979). Moral Hazard and Observability. *The Bell Journal of Economics*, 10(1), 74–91.
- Holmstrom, B., & Milgrom, P. (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica*, 55(2), 303–328.
- Holmstrom, B., & Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization*, 7, 24–52.
- Hull, C. L. (1932). The goal-gradient hypothesis and maze learning. *Psychological Review*, 39(1), 25–43.
- Hull, C. L. (1938). The goal-gradient hypothesis applied to some “Field-Force” problems in the behavior of young children. *Psychological Review*, 45(4), 271–298.
- Joseph K., & Kalwani, M. (1998). The role of bonus pay in sales force compensation plans. *Industrial Marketing Management*, 27(2), 147–159.
- Kahneman, D., & Tversky, A. (1979), Prospect theory: an analysis of decision under risk. *Econometrica*, 47(2), 263–292.
- Kim, S. K. (1997). Limited liability and bonus contracts. *Journal Economics & Management Strategy* 6(4):899–913.
- Kishore, S., Rao, R. S., Narasimhan, O., & John, G. (2013). Bonuses versus commissions: A field study. *Journal of Marketing Research*, 50(3), 317–333.
- Kivetz R., Urminsky, O., & Zheng, Y. (2006). The goal-gradient hypothesis resurrected: Purchase acceleration, illusionary goal progress, and customer retention. *Journal of Marketing Research*, 43(1), 39–58.
- Lal, R., & Srinivasan, V. (1993). Compensation plans for single- and multi-product sales forces: An application of the Holmstrom-Milgrom model. *Management Science*, 39(7), 777–793.
- Latham, G. P., & Locke, E. A. (1991). Self-regulation through goal setting. *Organizational Behavior and Human Decision Processes*, 50(2), 212–247.
- Lepper, M., Greene, D., & Nisbet, R. (1973). Undermining children’s intrinsic interest with extrinsic reward; a test of “overjustification” hypothesis. *Journal of Personality and Social Psychology* 28(1), 129–37.

Nunes, J. C., & Drèze, X. (2006). The endowed progress effect: How artificial advancement increases effort. *Journal of Consumer Research*, 32(4), 504–512.

Oyer, P. (1998). Fiscal year ends and nonlinear incentive contracts: The effect on business seasonality. *Quarterly Journal of Economics*, 113(1), 149–185.

Oyer, P. (2000). A theory of sales quotas with limited liability and rent sharing. *Journal of Labor Economics*, 18(3), 405–426.

Schweitzer, M. E., Ordóñez, L., & Douma, B. (2004). Goal setting as a motivator of unethical behavior. *Academy of Management Journal*, 47(3), 422–432.

U.S. Department of Labor, Bureau of Labor Statistics (2018). *National Occupational Employment and Wage Estimates*, May.

World Economic Forum (2014). *Global Gender Gap Report*.
<http://reports.weforum.org/global-gender-gap-report-2014/>

Zhang, Y., & Gao, L. (2016). Wanting ever more: Acquisition procedure motivates continued reward acquisition. *Journal of Consumer Research*, 43(2), 230–245.

Zoltners, A. A., Sinha, PK, & Lorimer, S. E. (2013). Are you paying enough attention to your sales force? *HBR Blog Network (blog)*, *Harvard Business Review*, April.

Zoltners, A. A., Lorimer, S. E., & Sinha, PK (2016). Wells Fargo and the slippery slope of sales incentives. *digital article*, *Harvard Business Review*, September.

Appendix A. Proof of Proposition 1

A.1. Proof of Proposition 1-i

First, we show that $\mu_1 \geq \mu_N$ for all N with the additional assumption $p^H = 0$. This assumption will be dropped shortly. Under this assumption: Observation 1) the agent never exerts effort on high-ticket products; thus, Observation 2) μ_N is never larger than $p^L \chi^L$ for any $N \geq 1$, where the upper bound $p^L \chi^L$ is achieved if and only if the agent exerts effort on low-ticket products in all periods.

Suppose that $N = 1$. If the agent exerts effort on a low-ticket product, $X = \chi^L$; and, thus, the agent attains compensation $\beta(\chi^L)\chi^L$ with probability p^L . The agent's compensation is zero with the complementary probability. It is optimal for the agent to exert effort on low-ticket products if and only if $p^L \beta(\chi^L)\chi^L \geq c^L$, and, thus,

$$\mu_1 = \begin{cases} p^L \chi^L & \text{if } p^L \beta(\chi^L)\chi^L \geq c^L, \\ 0 & \text{if } p^L \beta(\chi^L)\chi^L < c^L. \end{cases}$$

Next, we show that $\mu_1 \geq \mu_N$ for any $N > 1$. This is clearly true when $p^L \beta(\chi^L)\chi^L \geq c^L$ because μ_1 already achieves the upper bound $p^L \chi^L$. Hence, it suffices to show that $\mu_1 = \mu_N = 0$, so that the most frequent quota cycle $N = 1$ is vacuously optimal for the other case, $p^L \beta(\chi^L)\chi^L < c^L$. To show this, suppose, hypothetically, that the agent faces an alternative commission scheme $\tilde{\beta}$ such that

$$\tilde{\beta}(x) = \beta(\chi^L) \quad \forall x \geq 0.$$

We prove the following two claims; $\mu_1 = \mu_N = 0$ follow these two claims.

- (a) The agent exerts (weakly) more effort under scheme $\tilde{\beta}$ than under β for any $N \geq 1$.
- (b) The agent exerts no effort under $\tilde{\beta}$, so also does not under β .

To prove Claim (a), note first that β is monotone by assumption, and, thus, $\tilde{\beta}(x) \geq \beta(x)$ for any $x \in [0, \chi^L]$. Furthermore, the agent's average sales volume never exceeds χ^L , and, thus, $\tilde{\beta}(X/N) \geq \beta(X/N)$ for any realization of the agent's performance. This implies that the agent exerts (at least weakly) more effort under compensation scheme $\tilde{\beta}$ than under β , regardless of N ,

because the agent's return from additional effort is always (at least weakly) higher under $\tilde{\beta}$, while the the disutility from additional effort is identical under the two schemes.

Next, to prove Claim (b), note that the commission rate is constant regardless of an agent's performance under compensation scheme $\tilde{\beta}$, implying that the agent's current performance does not affect any future rewards, and, thus, the agent will act identically for any $N \geq 1$. Hence, it suffices to show that the agent never exerts effort when $N = 1$. If the agent exerts effort when the quota cycle is $N = 1$, his or her expected payoff will be

$$p^L \tilde{\beta}(\chi^L) \chi^L - c^L = p^L \beta(\chi^L) \chi^L - c^L < 0,$$

where the equality is due to the definition of $\tilde{\beta}$, and the inequality is due to the assumption that $p^L \beta(\chi^L) \chi^L < c^L$. Therefore, the agent never exerts effort under commission scheme $\tilde{\beta}$ for any $N \geq 1$, so also does not under commission scheme β .

Finally, the proof is completed by observing that the above arguments continue to be valid when p^H is positive but close to zero; the agent still never exerts effort on high-ticket products because doing so can never compensate for his or her disutility $c^H > 0$, and one can easily see that all of the above arguments remain valid.

A.2. Proof of Proposition 1-ii

Analogous to Section A.1, we first prove the proposition with the additional assumption $p^H = 0$. This assumption will be dropped shortly. If the firm chooses $N = 1$, the agent will choose a^L in every period because $u^L > 0 > u^H$: $u^L = p^L \chi^L \beta(\chi^L) - c^L > 0$ from the assumption given in Proposition 1-ii; and $u^H = -c^H < 0$ from the assumption $p^H = 0$. Hence, $\mu_1 = p^L \chi^L$.

Suppose that the firm chooses $N > 1$ and $\beta(\chi^L/N) < c^L/(p^L \chi^L) < \beta(\chi^L)$, as assumed in Proposition 1-ii. Note that the case of zero cumulative sales up to period $N-1$ (that is, $x_1 = x_2 = \dots = x_{N-1} = 0$ in all periods up to period $N-1$) is a positive-probability event. The agent will take the null action in the last period N conditional on this event. Clearly, the agent will never choose $a_N = a^H$ given that $p^H = 0$. If the agent takes $a_N = a^L$, he or she will generate $x_N = \chi^L$ and, thus, end up with $\bar{x} = \chi^L / N$ and $X = \chi^L$ with probability p^L ; and $\bar{x} = X = 0$ with the

complementary probability. Hence, the agent's expected continuation payoff from choosing $a_N = a^L$ (conditional on $x_1 = x_2 = \dots = x_{N-1} = 0$) is

$$p^L \beta(\chi^L / N) \chi^L - c^L.$$

Recall that $\beta(\chi^L / N) < c^L / (p^L \chi^L)$ by assumption, and, thus, the agent's expected payoff is negative:

$$p^L \beta(\chi^L / N) \chi^L - c^L < p^L \left[\frac{c^L}{p^L \chi^L} \right] \chi^L - c^L = c^L - c^L = 0.$$

The negative payoff is clearly suboptimal for the agent, as he or she can at least guarantee zero continuation payoff by taking the null action. Hence, the agent will give up and take the null action in the last period, conditional on $x_1 = x_2 = \dots = x_{N-1} = 0$.

The above observation implies that a positive-probability event will induce the agent to take the null action at some point within a quota cycle, and, thus, the firm's average sales volume will be strictly less than $\mu_4 = p^L \chi^L$. Finally, the proof is completed by observing that the above arguments continue to be valid when p^H is positive but close to zero; the agent still never exerts effort on high-ticket products because doing so can never compensate for his or her disutility $c^H > 0$, and one can easily check that all the arguments above remain valid.

Appendix B. Proof of Proposition 2

B.1 Reformulation of the Agent's Problem

For this appendix, define $\lambda^\theta = p^\theta \chi^\theta$ for both $\theta = H$ and L . Consider the following alternative but mathematically equivalent formulation of the agent's optimization problem. The agent works over a unit interval of time indexed by $t \in [0, 1]$. At each *discrete* time $t_k = \frac{k-1}{N}$, $k = 1, \dots, N$, the agent takes an action $a_{t_k} \in \{a^H, a^L, a^\varnothing\}$, in which case the agent incurs disutility $c(a_{t_k}) \in \{c^H, c^L, 0\}$, and sales volume $x_{t_k} \in \{\chi^H, \chi^L, 0\}$ may occur, as in the original model. For any realization of the total sales volume $X = \sum_{k=1}^N x_{t_k}$, the agent receives compensation $\beta(X / N) X$ at the end of the quota cycle.

Let

$$V_N(\sigma) = \mathbb{E} \left[\frac{\beta(X/N)X}{N} - \frac{\sum_{k=1}^N c(a_{t_k})}{N} \right]$$

denote the agent's normalized expected payoff (total payoff divided by N) under a policy σ .

B.2 Proof of Proposition 2

Let σ_N^* be the agent-optimal policy for each quota cycle N , so that $V_N(\sigma)$ is maximized at $\sigma = \sigma_N^*$.

If there are multiple agent-optimal policies, choose one that generates the highest expected sales volume for the firm. As the agent follows the optimal policy σ_N^* , the agent generates the following stochastic process:

$$z = (z_k)_{k=1}^N = (a_{t_k}, x_{t_k})_{k=1}^N.$$

Let Z_N denote the set of all feasible z and P_N denote the probability law over Z_N induced by the agent's optimal policy. For each $z \in Z_N$, let $V_N(z)$ denote the agent's total final payoff conditional on z being realized. Note that $V_N(z)$ is a random variable whose realization depends on the realization of z , and $V_N(\sigma_N^*) = \mathbb{E}^{P_N}[V_N(z)]$ denotes the agent's ex ante expected payoff from the agent-optimal policy.

Suppose that the agent focuses on high-ticket products over a time interval $[t, t + dt)$. The number of t_k in this interval is approximately Ndt ;²⁶ hence, the agent incurs disutility

$$\frac{\sum_{t_k \in [t, t+dt)} c^H}{N} = \frac{\sum_{t_k \in [t, t+dt)} c^H}{Ndt} dt \approx c^H dt,$$

where the error of the approximation vanishes as $N \rightarrow \infty$. Also, by the law of large numbers,

$$\frac{\sum_{t_k \in [t, t+dt)} x_{t_k}}{N} \text{ converges in probability to } \lim_{N \rightarrow \infty} \mathbb{E} \left[\frac{\sum_{t_k \in [t, t+dt)} x_{t_k}}{Ndt} \right] dt = \lambda^H dt$$

as $N \rightarrow \infty$. We can obtain similar results for the case in which the agent focuses on low-ticket products over a time interval $[t, t + dt)$, replacing all the superscripts H in the above equations with L .

²⁶ The exact number of t_k is the floor of Ndt (the smallest integer less than Ndt). The error of the approximation vanishes as N goes to infinity.

The above observation indicates that the law of large numbers eliminates the randomness in the agent's optimization problem in the limiting case, such that $N \rightarrow \infty$. The next lemma, which is key to the proof of Proposition 2, shows that the distribution of $V_N(z)$, indeed, converges to a constant distribution (degenerate distribution). Define $\tau^H := \frac{q-\lambda^L}{\lambda^H-\lambda^L} \in (0,1)$ and $U := u^H \tau^H + u^L (1 - \tau^H)$. U is strictly positive under Assumptions 3-A and 3-B.

Lemma B.1. There is a strictly increasing sequence of quota cycles $(N_n)_{n=1}^\infty$ such that $V_{N_n}(z)$ converges to U in probability;²⁷ that is, $\lim_{n \rightarrow \infty} P_{N_n} \{ |V_{N_n}(z) - U| < \varepsilon \} = 1$ for any $\varepsilon > 0$.

Proof. See Appendix C. ■

We now prove Proposition 2. Choose $\varepsilon > 0$ such that $U - \varepsilon > 0$, and construct a strictly increasing sequence of quota cycles $(N_n)_{n=1}^\infty$ such that $\lim_{n \rightarrow \infty} P_{N_n} \{ |V_{N_n}(z) - U| < \varepsilon \} = 1$. The agent's payoff $V_{N_n}(z)$ can be positive only if the quota is achieved (otherwise, the agent receives zero commission, which necessarily results in a non-positive total payoff), and, thus,

$$P_{N_n} \{ X / N \geq q \} \geq P_{N_n} \{ V_{N_n}(z) > U - \varepsilon > 0 \} \geq P_{N_n} \{ |V_{N_n}(z) - U| < \varepsilon \}$$

for any n , which implies that $P_{N_n} \{ X / N \geq q \}$ also converges to 1. Note that

$$\mu_{N_n} \geq q \cdot P_{N_n} \{ X / N \geq q \}$$

for any n , and the right-hand side converges to q . Hence, there must be a quota cycle N such that $\mu_N > \mu_1 = \lambda^L = p^L \chi^L$.

Appendix C. Proof of Lemma B.1

Suppose that the agent follows the optimal policy σ_N^* for all N . Also, define $\tau^H := \frac{q-\lambda^L}{\lambda^H-\lambda^L} \in (0,1)$ and $U := u^H \tau^H + u^L (1 - \tau^H)$ as in Section B.2. To prove Lemma B.1, we first make two preliminary observations (Lemmas C.1 and C.2), which together will imply Lemma B.1. The first observation is

²⁷ A sequence $(N_n)_{n=1}^\infty$ is strictly increasing if $N_{n+1} > N_n$ for any n .

that U is the asymptotic lower bound for the expected total payoff $V_N(\sigma_N^*) = \mathbb{E}[V_N(z)]$. Note that this bound coincides with the limit of $V_N(z)$ in Lemma B.1.

Lemma C.1. $\liminf_{N \rightarrow \infty} V_N(\sigma_N^*) \geq U$.

Proof. Fix a small number $\varepsilon > 0$ and consider the following policy $\sigma_N^b \in \Sigma_N$ for each N (which is not necessarily agent-optimal): $a_{t_k} = a^H$ at any time $t_k < \tau^H + \frac{2\varepsilon}{u^L - u^H}$ and $a_{t_k} = a^L$ at any time $t_k \geq \tau^H + \frac{2\varepsilon}{u^L - u^H}$. Utilizing policy σ_N^b , the agent incurs disutility (approximately) equal to

$$\frac{\sum_{k=1}^N c(a_{t_k})}{N} \approx c^H \left(\tau^H + \frac{2\varepsilon}{u^L - u^H} \right) + c^L \left(1 - \tau^H - \frac{2\varepsilon}{u^L - u^H} \right) = c^H \tau^H + c^L (1 - \tau^H) + 2\varepsilon \frac{c^H - c^L}{u^L - u^H}.$$

By the law of large numbers, the expected average sales volume is (approximately)

$$\mathbb{E} \left[\frac{\sum_{t_k \in [0,1]} x_{t_k}}{N} \right] \approx \lambda^H \tau^H + \lambda^L (1 - \tau^H) + 2\varepsilon \frac{\lambda^H - \lambda^L}{u^L - u^H} = q + 2\varepsilon \frac{\lambda^H - \lambda^L}{u^L - u^H},$$

and the quota q is achieved with probability approaching 1. Therefore,

$$\begin{aligned} \lim_{N \rightarrow \infty} V_N(\sigma_N^b) &\geq b \left[\lambda^H \tau^H + \lambda^L (1 - \tau^H) + 2\varepsilon \frac{\lambda^H - \lambda^L}{u^L - u^H} \right] - \tau^H c^H - (1 - \tau^H) c^L - 2\varepsilon \frac{c^H - c^L}{u^L - u^H} \\ &= u^H \tau^H + u^L (1 - \tau^H) - 2\varepsilon. \end{aligned}$$

Finally, Lemma C.1 follows the observation that $\liminf_{N \rightarrow \infty} V_N(\sigma_N^*) \geq \lim_{N \rightarrow \infty} V_N(\sigma_N^b)$ and $\varepsilon > 0$ is an arbitrarily chosen small number. \blacksquare

The next observation is that U also bounds $V_N(z)$ (not its expectation) *approximately* (with probability approaching 1). Note that the bound U coincides with the limit of $V_N(z)$ in Lemma B.1 and the asymptotic lower bound of $V_N(\sigma_N^*)$ in Lemma C.1.

Lemma C.2. There exist an increasing sequence $(N_n)_{n=1}^\infty$ of quota cycles and a positive number $\bar{\varepsilon} > 0$ such that $\lim_{n \rightarrow \infty} P_{N_n} \left\{ V_{N_n}(z) \leq U + \varepsilon \right\} = 1$ for any $\varepsilon \in (0, \bar{\varepsilon})$.

Before the proof of Lemma C.2, we first discuss how we prove Lemma B.1 with Lemmas C.1 and C.2. By Lemmas C.1 and C.2 and the basic property of the limit infimum, there is an increasing

sequence $(N_n)_{n=1}^\infty$ such that $\lim_{n \rightarrow \infty} V_{N_n}(\sigma_{N_n}^*) = \lim_{n \rightarrow \infty} \mathbb{E}[V_{N_n}(z)] = U$. Furthermore, utilizing Lemma C.2, we may assume that $\lim_{n \rightarrow \infty} P_{N_n} \{V_{N_n}(z) \leq U + \varepsilon\} = 1$ for all sufficiently small $\varepsilon > 0$. Next, we prove that $\text{Var}[V_{N_n}(z)] = \mathbb{E}[(V_{N_n}(z))^2] - (\mathbb{E}[V_{N_n}(z)])^2$ converges to zero as $n \rightarrow \infty$, and, thus, $V_{N_n}(z)$ converges to $\lim_{n \rightarrow \infty} \mathbb{E}[V_{N_n}(z)] = U$ in probability, as stated in Lemma B.1. Note that

$$\lim_{n \rightarrow \infty} \mathbb{E}[(V_{N_n}(z))^2] \leq \lim_{n \rightarrow \infty} P_{N_n} \{V_{N_n}(z) \leq U + \varepsilon\} \cdot (U + \varepsilon)^2 + \lim_{n \rightarrow \infty} (1 - P_{N_n} \{V_{N_n}(z) \leq U + \varepsilon\}) \cdot V^{\max},$$

where V^{\max} is the largest value that $V_{N_n}(z)$ can take. $P_{N_n} \{V_{N_n}(z) \leq U + \varepsilon\}$ converges to 1 and, thus,

$$\lim_{n \rightarrow \infty} \mathbb{E}[(V_{N_n}(z))^2] \leq (U + \varepsilon)^2. \text{ By construction, } \lim_{n \rightarrow \infty} (\mathbb{E}[V_{N_n}(z)])^2 = U^2 \text{ as } \lim_{n \rightarrow \infty} \mathbb{E}[V_{N_n}(z)] = U.$$

Combined,

$$\lim_{n \rightarrow \infty} \text{Var}[V_{N_n}(z)] = \lim_{n \rightarrow \infty} \mathbb{E}[(V_{N_n}(z))^2] - \lim_{n \rightarrow \infty} (\mathbb{E}[V_{N_n}(z)])^2 = (U + \varepsilon)^2 - U^2.$$

The conclusion that $\lim_{n \rightarrow \infty} \text{Var}[V_{N_n}(z)] = 0$ follows once we take the limit $\varepsilon \rightarrow 0$, which completes the proof of Lemma B.1.

Proof of Lemma C.2. In the remaining part of Appendix C, we prove Lemma C.2. As the agent follows the optimal policy σ_N^* , the agent generates the stochastic process $z = (z_{t_k})_{k=1}^N = (a_{t_k}, x_{t_k})_{k=1}^N$. For each realization z , let

$$z^H = \{z_{t_k} = (a_{t_k}, x_{t_k}) : a_{t_k} = a^H\}$$

be the collection of terms in z such that $a_{t_k} = a^H$, and define z^L similarly. Then, $|z^H|$ (respectively, $|z^L|$) is the number of periods in which the agent focuses on a high-ticket product (respectively, a low-ticket product). Finally, define for each N and $\varepsilon > 0$

$$Z_N^H(\varepsilon) = \left\{ z \in Z_N : \sum_{k: a_{t_k} = a^H} \frac{x_{t_k}}{N} \in \left[\frac{\lambda^H |z^H|}{N} - \frac{\varepsilon}{2}, \frac{\lambda^H |z^H|}{N} + \frac{\varepsilon}{2} \right] \text{ and } |z^H| \geq \varepsilon N \right\},$$

$$Z_N^L(\varepsilon) = \left\{ z \in Z_N : \sum_{k: a_{t_k} = a^L} \frac{x_{t_k}}{N} \in \left[\frac{\lambda^L |z^L|}{N} - \frac{\varepsilon}{2}, \frac{\lambda^L |z^L|}{N} + \frac{\varepsilon}{2} \right] \text{ and } |z^L| \geq \varepsilon N \right\}.$$

To understand which event $Z_N^H(\varepsilon)$ represents ($Z_N^L(\varepsilon)$ can be understood similarly), note first that the condition $|z^H| \geq \varepsilon N$ holds in the limiting case $N \rightarrow \infty$ only if $|z^H|$ also goes to infinity;

hence, this condition simply requires the agent to choose a^H arbitrarily many times in the limit. Next, provided that $|z^H|$ is sufficiently large, the law of large numbers implies that $\sum_{k:a_k=a^H} x_{t_k} / N$ lies in the interval $\left[\frac{\lambda^H |z^H|}{N} - \frac{\varepsilon}{2}, \frac{\lambda^H |z^H|}{N} + \frac{\varepsilon}{2} \right]$ with probability approaching 1. However, for any *finite* N , there is still a small probability that $\sum_{k:a_k=a^H} x_{t_k} / N$ deviates from this interval. $Z_N^H(\varepsilon)$ excludes all such realizations of z ; $Z_N^H(\varepsilon)$ includes only realizations of z along which the law of large numbers *does* come into effect.

The next observation (Claim C.1) is that U serves as an *approximate* upper bound for the agent's payoff $V_N(z)$, conditional on $z \in Z_N^H(\varepsilon) \cap Z_N^L(\varepsilon)$.

Claim C.1. There exist $M > 0$ and $\bar{\varepsilon} > 0$ such that $\sup\{V_N(z) : z \in Z_N^L(\varepsilon) \cap Z_N^H(\varepsilon)\} \leq U + M\varepsilon$ for any $\varepsilon \in (0, \bar{\varepsilon})$ and $N > 0$.

Proof of Claim C.1. Fix $N > 0$ and $\varepsilon \in (0, 1)$. By definition, for any $z \in Z_N^L(\varepsilon) \cap Z_N^H(\varepsilon)$, the average sales volume at the end of the quota cycle is less than $\frac{\lambda^H |z^H|}{N} + \frac{\lambda^L |z^L|}{N} + \varepsilon$, and

$$V_N(z) \leq \beta \left(\frac{\lambda^H |z^H|}{N} + \frac{\lambda^L |z^L|}{N} + \varepsilon \right) \left[\frac{\lambda^H |z^H|}{N} + \frac{\lambda^L |z^L|}{N} + \varepsilon \right] - \frac{c^H |z^H|}{N} - \frac{c^L |z^L|}{N}.$$

$\frac{|z^H|}{N} \in [0, 1]$, $\frac{|z^L|}{N} \in [0, 1]$, $\left(\frac{|z^H|}{N} + \frac{|z^L|}{N} \right) \in [0, 1]$ by construction. Hence, the right-hand side in the

last inequality is bounded by the value of the following optimization problem:

$$\max_{v, w, v+w \in [0, 1]} \beta(\lambda^H v + \lambda^L w + \varepsilon)[\lambda^H v + \lambda^L w] - c^H v - c^L w + b(\lambda^H + \varepsilon)\varepsilon.$$

Note that the objective function is piecewise-linear (because β is piecewise-linear) and all the constraints are linear. Its solution is

$$v = \frac{q - \lambda^L - \varepsilon}{\lambda^H - \lambda^L} = \tau^H - \frac{\varepsilon}{\lambda^H - \lambda^L} \quad \text{and} \quad w = \frac{\lambda^H - q + \varepsilon}{\lambda^H - \lambda^L} = 1 - \tau^H + \frac{\varepsilon}{\lambda^H - \lambda^L},$$

and the value of the objective function at this optimum is $U + \left[\frac{u^L - u^H}{\lambda^H - \lambda^L} + b(\lambda^H + \varepsilon) \right] \varepsilon$. In conclusion,

$V_N(z) \leq U + \left[\frac{u^L - u^H}{\lambda^H - \lambda^L} + b(\lambda^H + \varepsilon) \right] \varepsilon$ for any $z \in Z_N^H(\varepsilon) \cap Z_N^L(\varepsilon)$ and sufficiently small $\varepsilon > 0$; hence,

Claim C.1 holds true with $M = \frac{u^L - u^H}{\lambda^H - \lambda^L} + b(\lambda^H + 1) > 0$ and $\bar{\varepsilon} = 1$. ■

The next natural question is what happens if $z \notin Z_N^H(\varepsilon) \cap Z_N^L(\varepsilon)$. Note that the law of large numbers implies that such a realization occurs with probability approaching zero, and, consequently, we may disregard the case $z \notin Z_N^H(\varepsilon) \cap Z_N^L(\varepsilon)$ provided that N is sufficiently large. The next claims formalize this idea. Throughout the remaining part in this section, let $A_N^L(\varepsilon)$ and $A_N^H(\varepsilon)$ denote the events

$$A_N^H(\varepsilon) := \left\{ z \in Z_N : \sum_{k:a_k = a^H} \frac{x_{a_k}}{N} \in \left[\frac{\lambda^H |z^H|}{N} - \frac{\varepsilon}{2}, \frac{\lambda^H |z^H|}{N} + \frac{\varepsilon}{2} \right] \right\}$$

and

$$A_N^L(\varepsilon) := \left\{ z \in Z_N : \sum_{k:a_k = a^L} \frac{x_{a_k}}{N} \in \left[\frac{\lambda^L |z^L|}{N} - \frac{\varepsilon}{2}, \frac{\lambda^L |z^L|}{N} + \frac{\varepsilon}{2} \right] \right\},$$

respectively. Note that $z \in Z_N^H(\varepsilon)$ if and only if $z \in A_N^H(\varepsilon)$ and $|z^H| > \varepsilon N$; similarly, $z \in Z_N^L(\varepsilon)$ if and only if $z \in A_N^L(\varepsilon)$ and $|z^L| > \varepsilon N$.

Claim C.2. Both $P_N \{ |z^L| > \varepsilon N \text{ and } z \notin A_N^L(\varepsilon) \}$ and $P_N \{ |z^H| > \varepsilon N \text{ and } z \notin A_N^H(\varepsilon) \}$

converge to zero as $N \rightarrow \infty$. Also,

$$\limsup_{N \rightarrow \infty} P_N \{ z \in Z_N^H(\varepsilon) \cap Z_N^L(\varepsilon) \} = \limsup_{N \rightarrow \infty} P_N \{ \min\{ |z^L|, |z^H| \} > \varepsilon N \}.$$

Proof of Claim C.2. To prove that $P_N \{ |z^L| > \varepsilon N \text{ and } z \notin A_N^L(\varepsilon) \}$ vanishes as $N \rightarrow \infty$, note that

$|z^L|$ remains larger than εN for all sufficiently large integers N only if $|z^L|$ itself also grows to infinity. This brings the law of large numbers into force for the agent's performance on low-ticket products, so we may ignore the case in which $|z^L| > \varepsilon N$ but $z \notin A_N^L(\varepsilon)$ in the limit. The proof for $P_N \{ |z^H| > \varepsilon N \text{ and } z \notin A_N^H(\varepsilon) \} \rightarrow 0$ is similar. Finally, to show that

$$\limsup_{N \rightarrow \infty} P_N \left\{ z \in Z_N^H(\varepsilon) \cap Z_N^L(\varepsilon) \right\} = \limsup_{N \rightarrow \infty} P_N \left\{ \min\{|z^L|, |z^H|\} > \varepsilon N \right\},$$

note first that

$$\begin{aligned} & P_N \left\{ \min\{|z^L|, |z^H|\} > \varepsilon N \text{ and } z \notin A_N^H(\varepsilon) \cap A_N^L(\varepsilon) \right\} \\ & \leq P_N \left\{ |z^H| > \varepsilon N \text{ and } z \notin A_N^H(\varepsilon) \right\} + P_N \left\{ |z^L| > \varepsilon N \text{ and } z \notin A_N^L(\varepsilon) \right\}, \end{aligned}$$

where the right-hand side converges to zero by the previous observations. Therefore,

$$\limsup_{N \rightarrow \infty} P_N \left\{ \min\{|z^L|, |z^H|\} > \varepsilon N \text{ and } z \notin A_N^H(\varepsilon) \cap A_N^L(\varepsilon) \right\} = 0$$

and

$$\limsup_{N \rightarrow \infty} P_N \left\{ \min\{|z^L|, |z^H|\} > \varepsilon N \right\} = \limsup_{N \rightarrow \infty} P_N \left\{ \min\{|z^L|, |z^H|\} > \varepsilon N \text{ and } z \in A_N^H(\varepsilon) \cap A_N^L(\varepsilon) \right\}.$$

To complete the proof, observe that two events, $\left\{ \min\{|z^L|, |z^H|\} > \varepsilon N \text{ and } z \in A_N^H(\varepsilon) \cap A_N^L(\varepsilon) \right\}$ and $\left\{ z \in Z_N^H(\varepsilon) \cap Z_N^L(\varepsilon) \right\}$, are equivalent, and, thus, the limit suprema of their probabilities coincide. ■

Claim C.3. There exist a sequence of quota cycles $(N_n)_{n=1}^\infty$ and a sequence of positive real numbers $(\varepsilon_n)_{n=1}^\infty$ such that $N_n \rightarrow \infty$, $\varepsilon_n \rightarrow 0$, and $P_{N_n} \left\{ z \in Z_{N_n}^H(\varepsilon_n) \cap Z_{N_n}^L(\varepsilon_n) \right\} \rightarrow 1$ as $n \rightarrow \infty$.

Proof of Claim C.3. We prove this claim through three steps.

Step 1. We first show that, for all sufficiently small $\varepsilon > 0$,

$$P_N \left\{ |z^H| \leq \varepsilon N \text{ and } V_N(z) > 0 \right\} \rightarrow 0 \quad \text{as} \quad N \rightarrow \infty.$$

Note that the above probability is the sum of three terms:

- (i) $P_N \left\{ \max\{|z^H|, |z^L|\} \leq \varepsilon N \text{ and } V_N(z) > 0 \right\}$;
- (ii) $P_N \left\{ z \in A_N^L(\varepsilon), |z^H| \leq \varepsilon N < |z^L|, \text{ and } V_N(z) > 0 \right\}$; and
- (iii) $P_N \left\{ z \notin A_N^L(\varepsilon), |z^H| \leq \varepsilon N < |z^L|, \text{ and } V_N(z) > 0 \right\}$.

We show that all three terms are zero or converge to zero as $N \rightarrow \infty$. First, the term (iii) converges to zero because

$$P_N \left\{ z \notin A_N^L(\varepsilon), |z^H| \leq \varepsilon N < |z^L|, \text{ and } V_N(z) > 0 \right\} \leq P_N \left\{ z \notin A_N^L(\varepsilon) \text{ and } \varepsilon N < |z^L| \right\},$$

where $P_N \{z \notin A_N^L(\varepsilon) \text{ and } \varepsilon N < |z^L|\}$ converges to zero by Claim C.2. To show that term (i) is zero, consider z such that $\max\{|z^L|, |z^H|\} \leq \varepsilon N$. For such a realization of z , provided that $\varepsilon > 0$ is sufficiently small, the quota is never achieved because

$$\sum_k \frac{x_{t_k}}{N} = \sum_{k: a_{t_k} = a^L} \frac{x_{t_k}}{N} + \sum_{k: a_{t_k} = a^H} \frac{x_{t_k}}{N} \leq \frac{|z^L| \chi^L + |z^H| \chi^H}{N} \leq \chi^H \varepsilon + \chi^L \varepsilon \leq 2\chi^H \varepsilon < q,$$

and, thus, $V_N(z) \leq 0$. This shows that $P_N \{\max\{|z^H|, |z^L|\} \leq \varepsilon N \text{ and } V_N(z) > 0\} = 0$.

To show that term (ii) is also zero, consider $z \in A_N^L(\varepsilon)$ such that $|z^H| \leq \varepsilon N < |z^L|$. For such a realization of z ,

$$\sum_k \frac{x_{t_k}}{N} = \sum_{k: a_{t_k} = a^L} \frac{x_{t_k}}{N} + \sum_{k: a_{t_k} = a^H} \frac{x_{t_k}}{N} \leq \left[\frac{|z^L| \chi^L}{N} + \frac{\varepsilon}{2} \right] + \frac{|z^H| \chi^H}{N}.$$

The last term on the right-hand side is less than $\varepsilon \chi^H$ by assumption, and the first term is bounded from above by $\lambda^L + \varepsilon$ (because $0 \leq |z^L|/N \leq 1$ by construction); hence, again, the right-hand side, as a whole, is less than q and $V_N(z) \leq 0$ provided that $\varepsilon > 0$ is sufficiently small. This proves that term (ii) is also zero.

Step 2. Choose $\alpha \in (0, 1)$ such that $U\alpha > \frac{u^H + U}{2}$. This is possible because $U > u^H$. We claim that

$$P_N \{ |z^L| \leq \varepsilon N \text{ and } V_N(z) \geq \alpha U \} \rightarrow 0 \quad \text{as } N \rightarrow \infty.$$

Note that this probability is the sum of three terms:

- (i) $P_N \{ \max\{|z^L|, |z^H|\} \leq \varepsilon N \text{ and } V_N(z) \geq \alpha U \}$;
- (ii) $P_N \{ z \in A_N^L(\varepsilon), |z^L| \leq \varepsilon N < |z^H|, \text{ and } V_N(z) \geq \alpha U \}$; and
- (iii) $P_N \{ z \notin A_N^L(\varepsilon), |z^L| \leq \varepsilon N < |z^H|, \text{ and } V_N(z) \geq \alpha U \}$.

We show that all three terms are zero or converge to zero as $N \rightarrow \infty$. Because αU is strictly positive, terms (i) and (iii) are (weakly) smaller than $P_N \{ \max\{|z^L|, |z^H|\} \leq \varepsilon N \text{ and } V_N(z) > 0 \}$, and $P_N \{ z \notin A_N^L(\varepsilon), |z^L| \leq \varepsilon N < |z^H|, \text{ and } V_N(z) > 0 \}$ respectively. Note that, by the observation in Step 1, $P_N \{ \max\{|z^L|, |z^H|\} \leq \varepsilon N \text{ and } V_N(z) > 0 \} = 0$; hence, term (i) is also zero. We can also show

that $P_N \left\{ z \notin A_N^H(\varepsilon), |z^L| \leq \varepsilon N < |z^H|, \text{ and } V_N(z) > 0 \right\}$ converges to zero as $N \rightarrow \infty$ for the same reason that $P_N \left\{ z \notin A_N^L(\varepsilon), |z^H| \leq \varepsilon N < |z^L|, \text{ and } V_N(z) > 0 \right\} \rightarrow 0$ in Step 1; hence, term (iii) also converges to zero as $N \rightarrow \infty$. To see that term (ii) also converges to zero, consider a realization of $z \in A_N^H(\varepsilon)$ such that $|z^L| \leq \varepsilon N < |z^H|$. By definition of $A_N^H(\varepsilon)$,

$$\sum_k \frac{x_k}{N} = \sum_{k:a_k=a^L} \frac{x_k}{N} + \sum_{k:a_k=a^H} \frac{x_k}{N} \leq \frac{|z^L| \chi^L}{N} + \frac{|z^H| \lambda^H}{N} + \frac{\varepsilon}{2}.$$

The first term on the right-hand side is less than $\varepsilon \chi^L$. Hence, provided that ε is sufficiently small, there are effectively no sales of low-ticket products along such a realization of z , while the agent's performance on high-ticket products is dictated by the law of large numbers. This also means that $V_N(z)$ is asymptotically bounded by

$$\max_{v \in [0,1]} \beta(\lambda^H v) - c^H v + O(\varepsilon) \leq u^H + O(\varepsilon) < \alpha U,$$

where $O(\varepsilon)$ stands for a term that converges to zero as $\varepsilon \rightarrow 0$. In conclusion, term (ii) also vanishes as $N \rightarrow \infty$.

Step 3. Fix $\varepsilon > 0$ and let $(N_n(\varepsilon))_{n=1}^\infty \subset \mathbb{N}$ be a strictly increasing sequence of quota cycles such that $P_{N_n(\varepsilon)} \left\{ \min\{|z^L|, |z^H|\} > \varepsilon N_n(\varepsilon) \right\}$ is convergent as $n \rightarrow \infty$ (such a convergent subsequence is guaranteed to exist, as probabilities are bounded between zero and one). Combining Claim C.2 and the observations made in the previous steps,

$$P_{N_n(\varepsilon)} \left\{ \min\{|z^L|, |z^H|\} \leq \varepsilon N_n(\varepsilon) \text{ and } V_{N_n(\varepsilon)} \geq \alpha U \right\} \rightarrow 0 \text{ as } n \rightarrow \infty,$$

which also implies that

$$\limsup_{n \rightarrow \infty} \sum_{z: \min\{|z^L|, |z^H|\} \leq \varepsilon N_n(\varepsilon)} V_{N_n(\varepsilon)}(z) P_{N_n(\varepsilon)}(z) \leq \alpha U \cdot \lim_{n \rightarrow \infty} P_{N_n(\varepsilon)} \left\{ \min\{|z^L|, |z^H|\} \leq \varepsilon N_n(\varepsilon) \right\}.$$

Also, note that

$$\limsup_{n \rightarrow \infty} \sum_{z: \min\{|z^L|, |z^H|\} > \varepsilon N_n(\varepsilon)} V_{N_n(\varepsilon)}(z) P_{N_n(\varepsilon)}(z) \leq (U + M\varepsilon) \cdot \lim_{n \rightarrow \infty} P_{N_n(\varepsilon)} \left\{ \min\{|z^L|, |z^H|\} > \varepsilon N_n(\varepsilon) \right\}$$

by Claims C.1 and C.2. On the other hand,

$$\limsup_{n \rightarrow \infty} V_{N_n(\varepsilon)}^* \geq \liminf_{n \rightarrow \infty} V_{N_n(\varepsilon)}^* \geq U$$

by Lemma C.1, and

$$\limsup_{n \rightarrow \infty} V_{N_n(\varepsilon)}^* \leq \limsup_{n \rightarrow \infty} \sum_{z: \min\{|z^L|, |z^H|\} > \varepsilon N_n(\varepsilon)} V_{N_n(\varepsilon)}(z) P_{N_n(\varepsilon)}(z) + \limsup_{n \rightarrow \infty} \sum_{z: \min\{|z^L|, |z^H|\} \leq \varepsilon N_n(\varepsilon)} V_{N_n(\varepsilon)}(z) P_{N_n(\varepsilon)}(z)$$

by the basic property of the limit supremum. All of the observations combined lead to

$$U \leq (U + M\varepsilon) \cdot \lim_{n \rightarrow \infty} P_{N_n(\varepsilon)} \left\{ \min\{|z^L|, |z^H|\} > \varepsilon N_n(\varepsilon) \right\} + \alpha U \cdot \lim_{n \rightarrow \infty} P_{N_n(\varepsilon)} \left\{ \min\{|z^L|, |z^H|\} \leq \varepsilon N_n(\varepsilon) \right\}.$$

By rearranging the terms, we see that

$$\lim_{n \rightarrow \infty} P_{N_n(\varepsilon)} \left\{ z \in Z_{N_n(\varepsilon)}^H(\varepsilon) \cap Z_{N_n(\varepsilon)}^L(\varepsilon) \right\} = \lim_{n \rightarrow \infty} P_{N_n(\varepsilon)} \left\{ \min\{|z^L|, |z^H|\} > \varepsilon N_n(\varepsilon) \right\} \geq 1 - \frac{M\varepsilon}{(1-\alpha)U + M\varepsilon},$$

where the equality is due to Claim C.2. The right-hand side is larger than $1 - \frac{2M\varepsilon}{(1-\alpha)U}$ for any sufficiently small $\varepsilon > 0$. Consequently,

$$\Lambda(\varepsilon) := \left\{ N \in \mathbb{N} \mid P_N \left\{ z \in Z_N^H(\varepsilon) \cap Z_N^L(\varepsilon) \right\} \geq 1 - \frac{2M\varepsilon}{(1-\alpha)U} \right\}$$

is non-empty and contains infinitely many positive integers.

To complete the proof of Claim C.3, construct a decreasing sequence ε_n of positive real numbers such that $\varepsilon_n \downarrow 0$. For each n , pick $N_n \in \Lambda(\varepsilon_n)$. Because each $\Lambda(\varepsilon_n)$ contains infinitely many positive integers, we may pick N_n so that $N_n \rightarrow \infty$. Now note that

$$P_{N_n} \left\{ z \in Z_{N_n}^H(\varepsilon_n) \cap Z_{N_n}^L(\varepsilon_n) \right\} \geq 1 - \frac{2M\varepsilon_n}{(1-\alpha)U}$$

for all sufficiently large n . The right-hand side converges to 1, and so does the left-hand side. \blacksquare

Finally, we prove Lemma C.2. Let M and $\bar{\varepsilon}$ be positive numbers for which Lemma C.1 holds true, and pick an arbitrary small number $\varepsilon \in (0, \bar{\varepsilon})$. Construct $(N_n)_{n=1}^\infty$ and $(\varepsilon_n)_{n=1}^\infty$ so that Claim C.3 holds true. Note that $M\varepsilon_n < \varepsilon$ for all sufficiently large n .

$$P_{N_n} \left\{ V_{N_n}(z) \leq U + \varepsilon \right\} \geq P_{N_n} \left\{ V_{N_n}(z) \leq U + M\varepsilon_n \right\} \geq P_{N_n} \left\{ z \in Z_{N_n}^H(\varepsilon_n) \cap Z_{N_n}^L(\varepsilon_n) \right\}$$

for all sufficiently large n , where the first inequality is due to the fact that $M\varepsilon_n < \varepsilon$, and the second inequality is due to Claim C.1. The right-hand side $P_{N_n} \left\{ z \in Z_{N_n}^H(\varepsilon_n) \cap Z_{N_n}^L(\varepsilon_n) \right\}$ converges to 1 by Claim C.3, and, thus, we obtain Lemma C.2 by taking $n \rightarrow \infty$ on both sides of the above inequality.

Table 1. Theoretical Prediction: Product Focus and Revenue

Agent type / Category	Product focus	Revenue
Low type	Unchanged from low-ticket products	Increase (due to an increase in the quantity of products sold)
High type	Changed from high- to low-ticket products	Decrease (due to a change in the product focus)

The table depicts the theoretical prediction provided in Section 3 as a firm changes its compensation plan from a long quota-cycle to a short quota-cycle plan.

Table 2. The Variable Compensation Plan

Level	1	2	3	4	5
Quota (\$sales/hour)	140	180	200	235	250
Commission rate (%)	0.27	0.67	0.9	1.5	2.0

The figures with regard to quota and the commission rate are approximate for confidentiality reasons.

Table 3. Summary Statistics by Group in the Pre-Intervention Period (April)

	Control	Treatment
Mean	154.28	152.29
Standard deviation	59.36	59.70

The figures denote the mean and standard deviation of SPH per day across salespeople and time. The unit is U.S. dollars.

Table 4. Average Sales per Hour by Group across Periods

	April	May
Control	149.17	162.75
Treatment	149.06	163.96

The figure denotes the average SPH per day for the control and treatment groups. The average SPH per day is computed by summing up all sales and dividing them by total working hours for each of the two groups. The unit is U.S. dollars.

Table 5. The Effect of Quota Frequency on Sales Performance

Model 1 (Homogeneous)		Model 2 (Heterogeneous)	
Daily quota	-0.003 (0.031)	Daily quota—Q1	0.111 (0.039)
		Daily quota—Q2	0.020 (0.038)
		Daily quota—Q3	-0.038 (0.036)
		Daily quota—Q4	-0.084 (0.036)
Agent fixed effects	Yes	Agent fixed effects	Yes
Time fixed effects	Yes	Time fixed effects	Yes

Dependent variable: the logarithm of SPH per day. Heteroscedasticity-consistent (Eicker–Huber–White) standard errors shown. Significance (at the 0.05 level) in bold.

Table 6. Placebo Test

	Placebo week in April			
	week 1	week 2	week 3	week 4
Daily quota—Q1	0.036 (0.083)	0.015 (0.070)	-0.055 (0.073)	-0.072 (0.069)
Daily quota—Q2	0.038 (0.082)	0.088 (0.067)	-0.055 (0.069)	-0.060 (0.066)
Daily quota—Q3	0.020 (0.077)	0.039 (0.064)	-0.025 (0.066)	-0.041 (0.063)
Daily quota—Q4	-0.017 (0.075)	0.077 (0.062)	-0.002 (0.064)	-0.029 (0.062)
Agent fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes

The Placebo tests are conducted using pre-intervention (April) data in which four separate weeks are used as imaginary treatment periods. Dependent variable: the logarithm of SPH per day. Heteroscedasticity-consistent (Eicker–Huber–White) standard errors shown. Significance (at the 0.05 level) in bold.

Table 7. The Effect of Quota Frequency on Product Returns

Model 1 (Homogeneous)		Model 2 (Heterogeneous)	
Daily quota	-0.040 (0.110)	Daily quota—Q1	0.068 (0.130)
		Daily quota—Q2	-0.025 (0.122)
		Daily quota—Q3	-0.104 (0.120)
		Daily quota—Q4	-0.080 (0.116)
Agent fixed effects	Yes	Agent fixed effects	Yes
Time fixed effects	Yes	Time fixed effects	Yes

Dependent variable: the logarithm of returns-to-sales ratio per day. Heteroscedasticity-consistent (Eicker–Huber–White) standard errors shown. Significance (at the 0.05 level) in bold.

Table 8. The Effect of Quota Frequency on Quantity and Price

Quantile \ Dependent variable	# products sold per hour	Average price of products sold
Daily quota—Q1	0.128 (0.036)	-0.014 (0.019)
Daily quota—Q2	0.044 (0.035)	-0.024 (0.018)
Daily quota—Q3	-0.017 (0.034)	-0.023 (0.017)
Daily quota—Q4	-0.041 (0.034)	-0.047 (0.017)
Agent fixed effects	Yes	Yes
Time fixed effects	Yes	Yes

Dependent variable: first column: the logarithm of sales quantity per hour per day; second column: the logarithm of average price of products sold per day. Heteroscedasticity-consistent (Eicker–Huber–White) standard errors used. Significance (at the 0.05 level) in bold.

Figure 1. Types of Variable Compensation Plans with Quotas

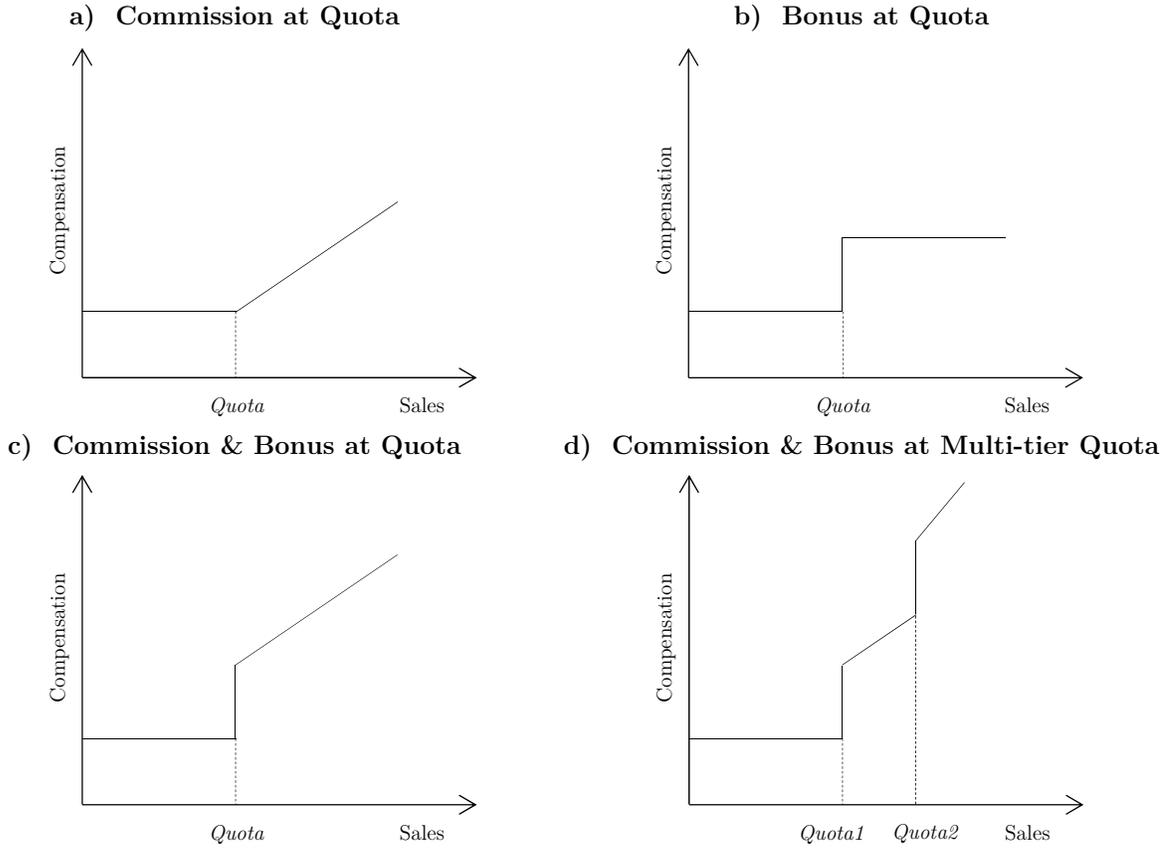
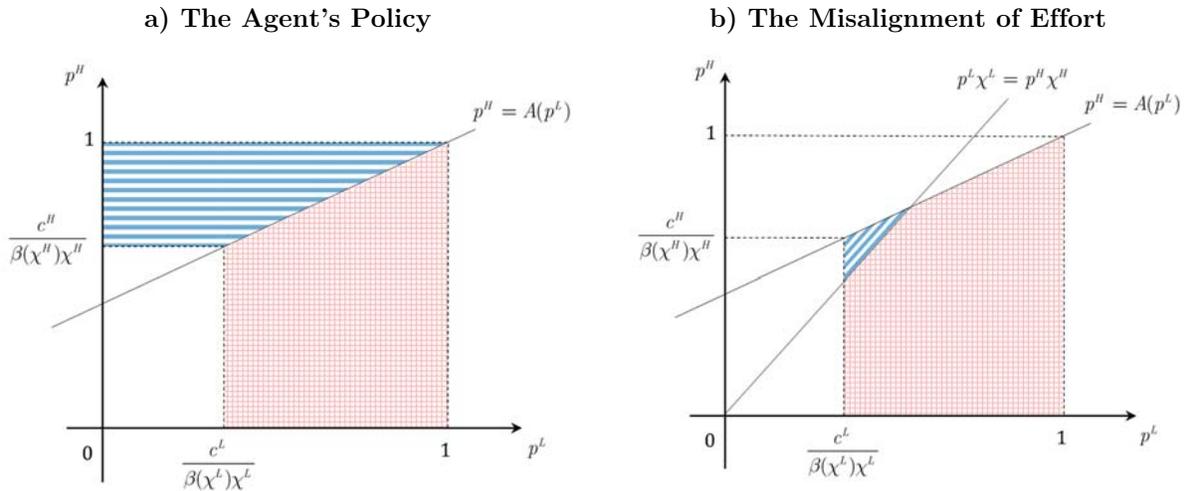


Figure 2. The Misalignment of the Agent's Effort with the Firm ($N=1$)

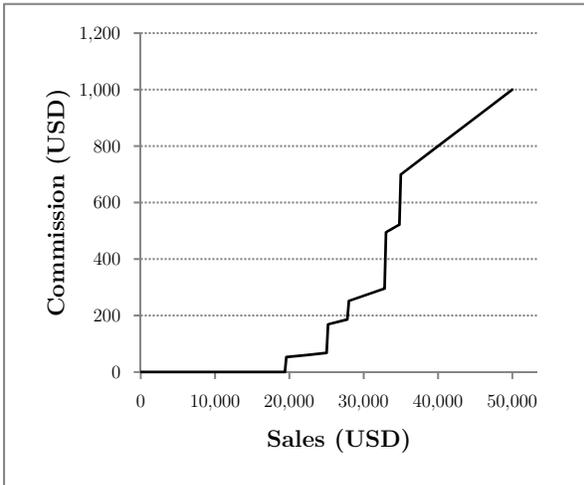


Horizontally shaded area: agents who focus on high-ticket products.
 Gridded area: agents who focus on low-ticket products.

Diagonally shaded area: agents who focus on low-ticket products but whose actions are suboptimal for the firm.
 Gridded area: agents who focus on low-ticket products and whose actions are optimal for the firm.

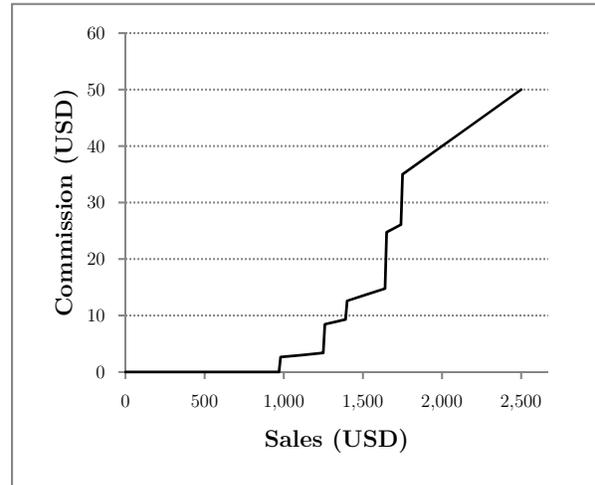
Figure 3. Relation between Sales and Commission

a) Monthly Quota Plan (April)



The figure illustrates monthly commission pay, conditional on sales, for a salesperson assigned 140 hours a month.

b) Daily Quota Plan (May)



The figure illustrates daily commission pay, conditional on sales, for a salesperson assigned 7 hours a day.