Self-Control and Incentives: An Analysis of Multiperiod Quota Plans

Sanjay Jain

Mays Business School, Texas A&M University, College Station, Texas 77843, sjain@tamu.edu

It is well known that individuals often fail to exert proper self-control. In organizational settings, this can lead to reduced productivity and profits. We use the literature on present-biased preferences to model employees' self-control problems and examine how firms can design compensation plans to reduce the negative consequences of their employees' self-control problems. Our results suggest that firms can mitigate self-control problems by delaying payment to the employees. This can be achieved by using multiperiod quotas (such as annual quotas) to compensate employees for their cumulative performance. Although such plans are prevalent in the market, there is little theoretical research that shows when multiperiod quota plans can be optimal. The paper provides one potential explanation for the widespread use of such quota plans. Interestingly, we find that such plans may be optimal despite the fact that they encourage more procrastination. We also find that such plans lead to higher effort by the employees and can sometimes improve the welfare of not only the firm but also the employees.

Key words: behavioral economics; game theory; present-biased preferences; compensation design
History: Received: June 27, 2011; accepted: February 12, 2012; Preyas Desai served as the editor-in-chief and Duncan Simester served as associate editor for this article. Published online in *Articles in Advance* May 22, 2012.

One of the greatest labor-saving inventions of today is tomorrow. —Vincent T. Foss

1. Introduction

Self-control problems are widespread among individuals. Failure to exert proper self-control can lead to various undesirable outcomes such as poor health, insufficient savings, addiction, poor productivity, and lost profits (e.g., Gruber and Köszegi 2001, O'Donoghue and Rabin 2000, Laibson 1997, Renn et al. 2011). There is now a substantial literature that has examined such self-control problems by appealing to the notion that consumers have present-biased preferences; i.e., their discount factor decreases over time (for empirical evidence of this phenomenon, see, for example, Thaler 1981, Chapman 1996, Kirby 1997, Benzion et al. 1989).[1] In an organization setting, such present-biased preferences can significantly impact performance. For example, consider a salesperson who has to decide whether to exert effort to close a sale or shirk. Although the salesperson knows that it is more beneficial for him to undertake the more effortful activity, when the time to exert the effort comes, he chooses to shirk or exert a suboptimal level of effort. This is because the discount factor for effort, which needs to be exerted immediately,

is much higher than the discount factor for the distant reward. This leads to lower sales for the firm and possibly lower compensation for the salesperson. Furthermore, because the salesperson makes effort decisions over multiple periods, the overall negative impact of self-control problems on firm's profits can be significant.

Some organizational behavior researchers have suggested that organizations and employees can benefit if the employees undergo self-management training (Frayne and Geringer 2000). Such training could potentially improve employees' performance and the firm's profits. However, the efficacy of these programs may be limited because the firm and the employees do not necessarily have the same goals. An alternative to this approach is to design incentive-compatible compensation plans that take into account the employees' self-control problems. Ideally, because the firm and the employees interact over multiple time periods, the firm would like to design compensation plans that are based on the employees' performance over multiple periods. Such an approach could improve firm's profits. Furthermore, because improved performance could lead to higher compensation, this approach could also benefit the employees. Despite the prevalence of self-control problems, there is almost no research that has examined how firms should design compensation plans in a multiperiod setting when employees have self-control

---

[1] There are alternative approaches for modeling self-control problems. See, for example, Thaler and Shefrin (1981), Gul and Pesendorfer (2001), and Fudenberg and Levine (2006).

problems.[2] The purpose of this paper is to develop an analytical model to address this issue. For illustrative purposes, we use the example of a salesperson and a firm, although the analysis is more generally applicable.

We follow the prior literature and model a salesperson's self-control problem by incorporating a quasi-hyperbolic discount parameter in the salesperson's utility function (see O'Donoghue and Rabin 1999a, b). We develop a four-period model. In period 0, the firm offers the salesperson a contract that he can accept or reject. Conditional on accepting the offer, the salesperson needs to decide on the amount of effort that he will exert in periods 1 and 2. The effort that the salesperson exerts results in sales. The salesperson can be paid only with a one-period delay after he exerts effort. This reflects the fact that effort does not lead to immediate sales, and sales are recorded with some delay. The problem for the firm is to design a compensation plan that considers the salesperson's present-biased preferences. The compensation plan needs to be accepted by the salesperson in period 0 and should provide sufficient incentives for him to implement the plan in periods 1 and 2.

We find several interesting results. We show that when salespersons have self-control problems, the firm needs to pay them more to induce them to exert more effort. Consequently, the salesperson is not necessarily hurt by an increase in self-control problems. Thus, the incentives for the salesperson to exert self-control may not be as strong as they are for the firm. Our results suggest that firms can mitigate the problem of shirking by delaying payments, such as by using a two-period quota. The use of such quotas is quite common in practice. For example, in their empirical study of salesforce compensation plans, Joseph and Kalwani (1998) find that 31% of firms in their sample use annual quotas, whereas another 23% use quarterly quotas. There is also some empirical research that suggests that the use of annual quotas (rather than only monthly quotas) can improve firms' profits (Chung et al. 2008). There is little (if any) theoretical research, however, that explains why firms use multiperiod bonuses and not pay bonuses more frequently. Thus, our results provide one potential explanation for the prevalence of multiperiod quotas.

We find that multiperiod quotas often lead to salespersons exerting more effort in both periods. The use of such quotas can, however, lead to effort distortion and to procrastination by the salesperson in the first period. Also, two-period quotas can sometimes lead to the salesperson decreasing effort in the second period, if he believes that he cannot make the

quota. Despite these negative consequences of quotas, we find that, in many cases, the use of such quota plans can significantly increase firm's profits. We also show that if the salesperson's self-control problem is not severe, then multiperiod quota plans decrease his long-run surplus. However, when the salesperson has a more severe self-control problem, then multiperiod quota plans can be a win–win for all parties: the salesperson, the firm, and the society is better off with such plans.

The remainder of this paper is organized as follows. In §2, we review the relevant literature. In §3, we present the base model analysis and extend the model in §4. In §5, we conclude our paper with managerial implications and directions for future research.

2. Related Literature

Our work is related to the growing literature that has used present-biased preferences to model self-control problems (e.g., Laibson 1997, Gruber and Köszegi 2001, Machado and Sinha 2007, Jain 2012). Present-biased preferences lead an individual to place greater weight on the present because his discount factor is decreasing over time. Such preferences are also time inconsistent. To see this, consider the case of an individual who is asked whether he would be willing to work for one hour on Monday or two hours on Wednesday for the same pay. If the question is asked a month before the work needs to be done, it is likely that the individual would prefer to work one hour on Monday. However, when it is Monday and the work needs to be performed immediately, the same individual may find it more attractive to enjoy the present and defer the work to Wednesday. Thus, present-biased preferences are time inconsistent and can lead to procrastination or shirking.

O'Donoghue and Rabin (1999b) examine how principals can design incentive plans for time-inconsistent agents when agents make a binary decision in each period whether to complete a task or defer completion. They consider incentive schemes to encourage the agents to complete tasks at the more efficient time. They find that deadline schemes with increasing penalties for noncompletion are optimal in such cases. Wu et al. (2009) study the design of compensation plans with present-biased agents in a team setting. In their framework, the firm has a task that needs to be completed in a specified amount of time, but the quality of the project is lower if agents procrastinate and defer the task to later periods. They consider a commission plan in which the firm pays the agent each period, depending on the amount of work completed. They show that if the impact of asymmetric efforts on quality is high, then the firm pays higher wages in the earlier periods. In contrast to this research, which is focused on the completion of one task, we examine

---

[2] In this paper, because we use present-biased preferences to model self-control problems, we will use these terms interchangeably.

the situation in which the agent decides *how much* effort to input in each period. Thus, in our framework, the agent decides on the amount of effort to exert, which, in turn, determines the amount of output.

Gilpatric (2008) studies design of contracts when agents have present-biased preferences. Similar to our paper, he examines the case where the agent decides on the quantity of effort that he will exert. Gilpatric considers the case when there is a mix of time-consistent and present-biased agents. His focus and results are substantially different from ours. In particular, he looks at the case when the agent exerts effort only once and the contract is therefore based on the results of a single period. His results suggest that if the firm has limited ability to punish agents and the agents have rational expectations, then the firm will offer a contract that will screen out present-biased agents. In contrast, if the agents are naïve and do not foresee their self-control problems, then the optimal contract accommodates shirking. In contrast to Gilpatric's (2008) approach, we look at multiperiod contracts. We find that when we consider the possibility that the agent exerts effort multiple times, offering multiperiod contracts can reduce the level of shirking. Furthermore, we also establish conditions under which the optimal compensation plan can increase not only firms' profits but also the agent's long-run surplus.[3]

Our work is also related to the vast literature on salesforce compensation design using the principal-agent framework (e.g., Basu et al. 1985, Rao 1990, Lal and Srinivasan 1993). In particular, our work is related to the literature that has studied the use of quota-bonus plans. Raju and Srinivasan (1996) show that, though not theoretically optimal, a simple compensation structure of salary plus linear commissions for all sales over quota may provide a reasonable approximation for the optimal plan. Oyer (2000) shows that bonus plans can be optimal when agents have limited liability and the participation constraint does not bind. Dai and Jerath (2011) find that quota plans can be optimal when we consider inventory constraints. This research has, however, not considered present-biased agents. Furthermore, there

is almost no theoretical research that considers the case of multiperiod quota plans and the reason for their prevalence. Our paper adds to this literature and shows that such plans can be optimal when salespersons are present biased.

This paper is more broadly related to the growing literature in marketing that tries to enrich standard economic models by incorporating psychological and sociological realism in these models (see, for example, Carpenter and Nakamoto 1990, Wernerfelt 1995, Amaldoss and Jain 2005, Syam et al. 2008, Villas-Boas 2009).

3. Model

We will consider the case of a salesperson and a firm. The firm wants the salesperson to exert effort in order to achieve sales. Exerting effort is costly for the salesperson, and the firm must compensate the salesperson for the effort. First, consider the one-period case, in which the salesperson exerts effort e at a cost $h(e)$, $h(0) = 0$, $h'(e) > 0$, and $h''(e) > 0$.[4] In other words, costs are increasing and convex in effort. These assumptions are standard. If the salesperson exerts effort e, then this leads to a sale of e units and total profits of $m \cdot e$, where m is the margin from the sale of each unit. The assumption that effort and sales are linearly related is without much loss in generality. We could assume that sales are a concave function of effort. This would then allow for the possibility that the salesperson is more likely to get higher sales if he smooths out effort over two periods, rather than concentrate his effort in one period.[5] To ensure that the firm would want the salesperson to exert effort, we assume that $\lim_{e \downarrow 0} h'(e) < m$.[6] Note that we are assuming that effort is deterministically related to sales. This may be appropriate in many contexts, where firms may be able to closely

---

[3] Another stream of research that is tangentially related to ours is one that focuses on the design of contracts for present-biased consumers. For example, DellaVigna and Malmendier (2004) study contract design in consumer markets and show that the use of two-part tariffs can be optimal. They show that when consumers are aware of their self-control problems, can then the firm benefit by offering a two-part tariff with high fixed cost and a per-unit price that is below marginal cost. Another related paper is Jain (2009), which studies how consumers can set goals to mitigate their self-control problems and maximize their own welfare. In contrast, this paper examines a principal–agent problem in which the objective is to maximize the principal's profits, even when it sometimes leads to lower surplus for the agent.

[4] Our framework can also be modified to allow for the possibility that there is a discontinuity at $e = 0$ and the initial effort requires a fixed setup cost f; i.e.,

$$\lim_{e \downarrow 0} h(e) = f > 0.$$

For small values of f, when the effort in each period is positive, this alternate formulation would not affect the main results. If f is large, then the salesperson might find it optimal to exert effort in only one period and therefore incur f only once. This would turn our two-period problem into a single-period one in which the salesperson exerts effort only in the last period. Even in this case, our proposed multiperiod contract would be more profitable than the single-period contract.

[5] To see this, assume that effort e leads to a sales $\phi(e)$, where $\phi(\cdot)$ is a concave function. Then the problem can be equivalently stated by defining $\tilde{e} = \phi(e)$, and the related cost function becomes $h(\phi^{-1}(\tilde{e}))$, where $\phi^{-1}(\cdot)$ is convex. Since $h(\cdot)$ is nondecreasing and convex, it follows that the redefined cost function $h \circ \phi$ still is convex in $\tilde{e}$.

[6] Note that if $h(\cdot)$ is of the form $h(e) = \eta e^{\lambda}$, then this condition is always satisfied for any $m > 0$.

monitor effort. In §§4.2 and 4.3, we consider situations in which the firm cannot infer effort from sales. This sequential analysis helps us better delineate the role of self-control problems and uncertainty on compensation plans.

Consider the case when the firm pays the salesperson an amount $R(e)$ based on sales e. We assume that the firm cannot pay the salesperson immediately but must wait some time before making the payment. More precisely, we assume that if the salesperson makes effort at time t, then sales materialize at time $t + 1$, which is the earliest time that the firm can pay the salesperson (see O'Donoghue and Rabin 1999a, and Gilpatric 2008 for a similar assumption). This assumption reflects the fact that effort rarely leads to immediate sales, and furthermore, typically, there is also delay in recording sales. We assume that the per-period (exponential) discount factor is 1. This is reasonable in most contexts. For example, consider the case of a salesperson who is paid at the end of the month. In this case, using the commonly used annual discount rate of 5%, the monthly discount factor is 0.996, which is close to 1. However, in §4.4, we also consider the case when the exponential discount factor is strictly less than 1.

3.1. Benchmark Case: No Self-Control Problems
First, consider the case when salespersons do not have self-control problems and have time-consistent preferences. The optimal plan is a nonlinear plan. Consider the case when the firm offers a payment $R(e)$ for output e.[7] We will restrict our analysis to the case where $R(e) \geq 0$. In other words, we rule out the possibility that the firm can impose penalties if the salesperson does not produce sufficient output. The worst that the firm can do is to not pay the salesperson. This is reasonable in most real-world contexts. Note that if we allow the firm to specify large penalties if the contracted sales are not achieved, then it is always possible to achieve first best outcomes.[8] Given a reward schedule $R(e)$, the salesperson will choose the effort, which maximizes $R(e) - h(e)$ subject to the constraint that $R(e) \geq h(e)$. Suppose, given an $R(e)$, the salesperson exerts an effort $e(R(e))$. The firm's problem is then to choose the optimal schedule $R(e)$ subject to the constraint that at the desired optimal effort e_n^*, $R(e_n^*) - h(e(R(e_n^*))) \geq 0$. The firm will want to pay the minimum amount $h(e_n^*)$ in order for the salesperson

to exert this effort. It remains to find the optimal e_n^*, which is given by

$$e_n^* = \arg \max_e [m \cdot e - h(e)]. \tag{1}$$

Since $\lim_{e \downarrow 0} h'(e) < m$, an optimal e_n^* exists and is unique because of the convexity of $h(\cdot)$. It is given by

$$e_n^* = h'^{-1}(m) \equiv \psi(m). \tag{2}$$

The nonlinear plan, in this case, can be implemented using a quota plan with a quota of $\psi(m)$ and a bonus of $h(\psi(m))$, which is given only if the quota is met. Now, consider the case when the salesperson needs to exert effort in two periods and is paid at the end of every period. In this case, the nonlinear plan that we just described continues to be optimal. In particular, the firm still continues to offer the quota plan that we derived and to offer bonuses each period.

3.2. Salespersons Have Self-Control Problems
Now consider the case when salespersons have self-control problems. We follow previous research, which has modeled self-control problems using the quasi-hyperbolic discount parameter (see, for example, O'Donoghue and Rabin 1999a, b). In particular, the discount function at time τ is given by

$$D(\tau) = \begin{cases} 1 & \text{if } \tau = 0, \\ \beta & \text{otherwise,} \end{cases} \tag{3}$$

where β is the quasi-hyperbolic discounting parameter where $0 < \beta < 1$.[9] Note that in this formulation, the salesperson's discounting depends on the time at which he makes the decision. As a consequence of this, the parameter β creates a time-inconsistency problem in that the salesperson's preferences change over time. To see this, consider the case when the salesperson has to make a decision whether to accept a contract in which he would be required to exert effort e in time period t, and the associated reward is R, which is given in time period $t + 1$. At time period 0, he will be willing to accept the contract as long as $R > h(e)$. However, at time period t when he has to immediately exert effort e, he would do so only if $\beta R > h(e)$. Thus, for $R \in (h(e), h(e)/\beta)$, the salesperson will shirk at time t, even though from his perspective at time 0, he should exert effort e at time period t. Thus, there is a discrepancy between the preferences of the salesperson in time period 0 and time period t.[10]

---

[7] In the deterministic case, effort e and output are directly related, which may not be the case when output is stochastically related to effort. We will discuss this case in §4.3.

[8] In our context, however, in the base model, there is no uncertainty, and if salespersons do not have self-control problems, then the first-best outcome can be achieved in any case. However, if the salespersons do have self-control problems, then the first-best outcome cannot be achieved when we assume limited liability.

[9] For empirical support for the existence of hyperbolic discounting in the salesforce context, see Chung et al. (2008).

[10] If the salesperson could be paid at the same time that he exerts effort, then the firm can induce higher levels of efforts with lower pay, and self-control issues become irrelevant. However, as we discussed before, such a concurrent payment is not a feasible option in many cases.

Note that the rational salesperson foresees that he is likely to have self-control problems at time t. We make the usual assumption that the salesperson has rational expectations. In §4.1, we will relax this assumption and study its consequences. With the assumption of rational expectations and quasi-hyperbolic discounting, the salesperson's decision is more appropriately modeled as a multiperson game in which the salesperson in each period acts as a Stackelberg leader, who fully takes into account the actions of the followers, i.e., his own actions in subsequent periods.

First, consider the case of nonlinear contracts where the salesperson is paid at the end of each period on the basis of his performance in that period. As before, in this case, the salesperson will accept an offer to exert effort e only if the reward $R \geq h(e)/\beta$. For it to be profitable for the firm to induce the salesperson to exert effort, we need a stronger condition than before; i.e., $\lim_{e \downarrow 0} h'(e) < m\beta$. The firm's problem is to choose the optimal effort e_s^* for this single-period plan while still satisfying the participation constraint that $R(e_s^*) \geq h(e_s^*)/\beta$. Because the firm would like to ensure that the participation constraint is just satisfied, the firm's problem is to choose

$$e_s^* = \arg\max_e \left[me - \frac{h(e)}{\beta} \right]. \tag{4}$$

The optimal effort that the firm would like to induce each period with such a plan is $e_s^* = h'^{-1}(m\beta) = \psi(m\beta)$. Since $h(\cdot)$ is increasing and convex, $\psi(\cdot)$ is clearly increasing. Thus, as expected, the salesperson's effort decreases as the self-control problem increases, i.e., β decreases.

Now let us consider the salesperson's surplus. In models with time-inconsistent preferences, analysis of surplus and social welfare can be problematic, because the preferences of the participants change over time (see Bernheim and Rangel 2005 for a discussion). Most papers in the literature have used the agent's long-run preferences as the appropriate measure of welfare (see, for example, Gruber and Köszegi 2001, O'Donoghue and Rabin 2003, DellaVigna and Malmendier 2004, Gilpatric 2008, Jain 2012). One way this can be justified is by asking the discount factor one would use to advice consumers (Harris and Laibson 2002). This approach also reflects the view that the hyperbolic discounting represents the consumer's self-control problem and not his "true" preferences (Akerlof 1991). We will use this approach to measure welfare.[11] With this approach,

the salesperson's long-run surplus in two periods, using single-period quota plans, is given by

$$U_0 = 2 \left[\frac{h(\psi(\beta m))}{\beta} - h(\psi(\beta m)) \right]$$
$$= 2 \left(\frac{1-\beta}{\beta} \right) h(\psi(\beta m)) > 0. \tag{5}$$

Note that $U_0 > 0$. This is in contrast to the case when the salesperson did not suffer from self-control problems and $\beta = 1$. Thus, from a long-run perspective, a salesperson's self-control problems make him better off. The result thus shows that self-help programs, which try to motivate salespersons to exert more self-control, may not always be very effective because the firm's and the salesperson's incentives are not necessarily aligned.

The firm's profits in this case are

$$\Pi = 2 \left[m\psi(\beta m) - \frac{h(\psi(\beta m))}{\beta} \right], \tag{6}$$

which are increasing in β. The total long-run welfare is given by $2m\psi(\beta m)$, which is also increasing in β. Thus, present-biased preferences make the salesperson better off, whereas the firm and society is worse off. To understand the intuition, note that when the salesperson has present-biased preferences, he can credibly commit to not putting in the required effort at times 1 and 2. The firm therefore must compensate the salesperson to provide enough incentives at times 1 and 2 to exert effort. This, however, makes the salesperson at time 0s, who views effort and reward equally, have strictly positive surplus.[12] We are interested in seeing whether we could devise a quota plan that would partly counteract the salesperson's advantage and increase the firm's profits. Interestingly, we find that delaying payment may be better than paying the salesperson at the end of each period, as the following proposition shows.[13]

PROPOSITION 1. *There always exists a (multiperiod) quota plan in which the firm pays $R_m^*(e_1, e_2)$, depending on the observed sales e_1 and e_2 in periods 1 and 2, such*

---

[11] Some authors have, however, argued for a more literal interpretation of the multiself model. Thus, we could assume that welfare improves only when the agent is better off in each period (see, for

example, Laibson 1997). This Pareto criterion, however, often leads to an incomplete ranking of allocations. Alternatively, we could use a weighted sum of the welfare of the agent in each period (see Bernheim and Rangel 2005).

[12] Because sales are deterministically related to effort, in this case, one may wonder why the firm cannot fire the salespersons who shirk. If, however, the pool of potential hires also have self-control problems (which is likely given the prevalence of these problems), such an approach does not help the firm much.

[13] Proofs of all propositions are in the technical appendix (http://mktsci.journal.informs.org/).

that the profits with the plan $R_m^*(e_1, e_2)$ are strictly higher than any single-period contract, as long as $\beta < 1$, where

$$R_m^*(e_1, e_2) = \begin{cases} \mathcal{B} & \text{if } e_1 + e_2 \geq \mathcal{Q}, \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

Figure 1 shows the optimal plan. The result therefore suggests that when the salesperson has self-control problems, the optimal solution may be to delay rewards by assigning him long-term goals. This seems counterintuitive. After all, the source of the salesperson's self-control problem is that he weighs the present much more than the future and is too short-term focused. To understand this result, consider the salesperson's problem. First, note that given our assumption of limited liability and hyperbolic discounting, the period 0 participation constraint never binds and the salesperson needs to make decisions in only two periods, i.e., periods 1 and 2. In the subsequent analysis, we will therefore not discuss period 0 constraints. With this, the salesperson will accept a contract $(\mathcal{Q}, \mathcal{B})$ if there exist effort levels (e_1, e_2) in periods 1 and 2 such that

$$\mathcal{Q} \leq e_1 + e_2, \quad (8)$$

$$\beta \mathcal{B} \geq h(e_1) + \beta h(e_2), \quad (9)$$

$$\beta \mathcal{B} \geq h(e_2), \quad (10)$$

where Equation (8) ensures that the quota is achieved with the effort profile (e_1, e_2), Equation (9) ensures that the salesperson exerts effort e_1 at time 1, and Equation (10) ensures that the salesperson exerts e_2 at time 2.

Consider the case when the firm sets the quota at $2\psi(\beta m)$. Recall that this is the total amount of effort that the firm will induce in two single-period contracts. Also, assume that the firm offers $\mathcal{B} =$

Figure 1 **Optimal Compensation Plan When Salespersons Have Present-Biased Preferences**

$(2h(\psi(\beta m)))/\beta - \epsilon$, where $\epsilon > 0$. In other words, the firm offers a contract that pays the salesperson a little less for the same amount of effort it can induce using two single-period contracts. First, consider (10). Because the bonus amount is cumulative, for a given level of effort in two periods, the salesperson is more motivated to exert effort e_2 in period 2. Thus, accumulation can enable us to reduce the total bonus slightly and still satisfy (10). Now consider the salesperson at time 1. From (9), we see that in period 1, this salesperson still values immediate effort more than the reward. However, he values both the effort and the reward equally when the effort needs to be exerted in period 2. This time inconsistency is what enables the firm to achieve the same level of effort in two periods by paying a cumulative amount, which is less than what it would pay under any single-period contract. It is also useful to note that absent time inconsistency, i.e., when $\beta = 1$, the multiperiod contract does not improve profits in our framework.

This result formalizes the general intuition that multiperiod bonuses (such as yearly bonuses) can increase motivation relative to other compensation approaches. There is discussion in the literature about the motivating role of bonuses, but it is not immediately clear why bonuses at the year's end should be any more motivational than bonuses each period.

Of course, the two-period bonus plan will not typically lead to the same efforts that would be exerted under the one-period plan, and the firm could do better by either inducing more effort or paying lower compensation or both. To address this, we now solve for the optimal plan. Given a quota plan, the salesperson will try to choose (e_1, e_2) to maximize

$$U_1 = \max_{e_1} \left\{ \beta \mathcal{B} - [h(e_1) + \beta h(\mathcal{Q} - e_1)] \right\}, \quad (11)$$

subject to the constraint (10). The following lemma establishes the conditions under which (10) binds for a particular functional form of $h(\cdot)$.

LEMMA 1. *Suppose $h(e) = \eta e^\lambda$, where η and λ are constants with $\eta > 0$ and $\lambda > 1$. Then there exists a β^* such that (10) binds if $\beta < \beta^*$ and does not bind otherwise, where β^* is implicitly defined by the equation*

$$1 - \beta - \beta^{\lambda/(\lambda-1)} = 0. \quad (12)$$

The optimal β^* would depend on the steepness of the effort function. Thus, if $\lambda = 2$, i.e., the effort function is quadratic, $\beta^* = 0.618$, while it is 0.569 if $\lambda = 3$. In other words, as λ increases, the range for which the constraint in (10) does not bind increases.

The firm will choose the optimal plan $(\mathcal{Q}, \mathcal{B})$ to maximize its profits, which are given by $m \cdot \mathcal{Q} - \mathcal{B}$. We have the following result.
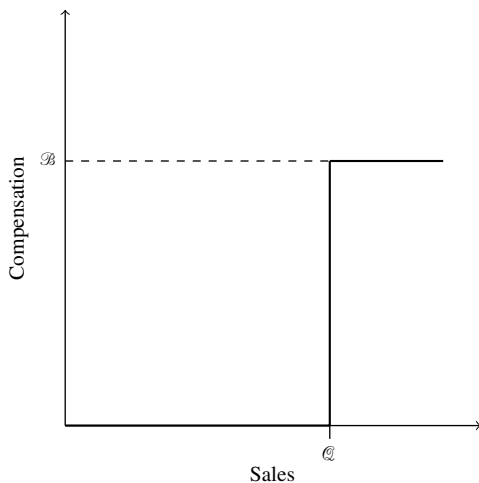
Proposition 2. *If under the optimal plan* (10) *does not bind, then*

a. *The optimal quota is* $\mathcal{Q} = \psi(m) + \psi(\beta m)$, *and* $\mathcal{B} = h(\psi(\beta m))/\beta + h(\psi(m))$.

b. *The salesperson exerts* $\psi(\beta m)$ *in period 1 and* $\psi(m)$ *in period 2. The overall effort that the salesperson exerts is higher and his long-run surplus is lower under the two-period quota plan than under the best single-period contract.*

Proposition 2 shows that when (10) does not bind, the firm can implement a plan in which the salesperson exerts $\psi(m)$ in period 2. Note that $\psi(m)$ is the optimal effort that the firm would like the salesperson to exert, if the salesperson did not have a self-control problem. Therefore, the result shows that the firm can partly counteract the adverse effect of β by combining and delaying rewards. Consistent with the empirical findings of Steenburgh (2008) and Chung et al. (2008), our results show that quota plans can lead to higher total effort. Note that unlike the case of single-period contracts, the salesperson does not exert the same level of effort in both periods. In fact, in period 1, he defers more work to be done in period 2. In other words, quotas lead to distortion of efforts and encourage procrastination in the first period. This effect has also been previously noted by Jain (2009) in the context of consumer decision making.[14] The result is also consistent with the empirical observation that sales tend to increase toward the end of the quota period (see, for example, Oyer 1998, Misra and Nair 2011). Intuition would suggest that these effort distortions reduce profits. However, our results show that the firm can benefit by offering such quota plans despite these effort distortions.

The proposition also shows that although multiperiod quotas help the firm, the salesperson is worse off under the multiperiod quota plan. To understand this result, note that the multiperiod quota plan affects the long-run surplus of the salesperson in two ways. First, it leads to higher effort being exerted by the salesperson. This aspect hurts the surplus that the salesperson receives. Second, the additional effort may also be accompanied by increased compensation. This aspect improves the salesperson's long-run surplus. Under the multiperiod quota plan, the firm is able to exploit the salesperson's time-inconsistent preferences. When (10) does not bind, this enables the firm to pay the salesperson less than the corresponding increase in his effort. Consequently, the

salesperson's long-run surplus goes down in this case. The following example illustrates the results from Proposition 2.

Example. Consider the case when $h(e) = e^2/2$. Recall that from Lemma 1, we know that in this case, (10) does not bind as long as $\beta > 0.618$. In this situation, under the single-period quota plan, the salesperson exerts $m\beta$ in each period and gets paid $(m^2\beta)/2$. The firm makes a total profit of $m^2\beta$ over two periods. In contrast, for the multiperiod quota case, the firm sets a quota of $m(1 + \beta)$ and offers a bonus of $(m^2(1+\beta))/2$. The salesperson exerts effort $m\beta$ in period 1 and m in period 2. The profits of the firm in this case are $(m^2(1+\beta))/2$. Thus, if $\beta = 0.7$, then the total sales and firm's profits improve by 21% if the firm employs multiperiod quotas as opposed to the case when it uses a single-period plan. Note that the firm pays a higher total compensation in the multiperiod quota case than under the single-period plan. Nevertheless, the salesperson is paid less corresponding to the increased effort. The salesperson's long-run surplus under the multiperiod plan is $(m^2(1-\beta)\beta)/2$, which is half his long-run surplus of $m^2(1-\beta)\beta$ under the single-period plan.

Now, consider the case when (10) binds. In this case, we have the following result.

Proposition 3. *Assume* (10) *binds. Then,*

a. *The optimal quota plan is such that the salesperson exerts effort* $e_2^c > e_1^c > \psi(m\beta)$.

b. *If* $h(e) = \eta e^\lambda$, *where* η *and* λ *are constants with* $\lambda > 1$, *then there exists a* $\hat{\beta} < \beta^*$ *such that the salesperson is better off with the optimal multiperiod quota plan than the optimal single-period contract when* $\beta < \hat{\beta}$.

Proposition 3 shows that even when (10) binds, under the optimal two-period plan, the salesperson exerts more effort than the one-period plan. Furthermore, the salesperson exerts less effort in the first period compared to the second period. Thus, the procrastination effect still holds even when (10) binds. Despite the procrastination effect, note that when (10) binds, the salesperson exerts more effort in *both* periods than he would in a single-period plan. This is unlike the case where (10) does not bind in which case the multiperiod plan only affects the second-period effort level. From Lemma 1, we know that for $h(e) = \eta e^\lambda$, (10) binds for small values of β. Therefore, in this situation, it implies that when the salesperson has a high level of self-control problems, the multiperiod quota plan will push the salesperson to exert more effort in both periods. The reason is that in this case, the salesperson in period 1 cannot implement his optimal plan to minimize $h(e_1) + \beta h(\mathcal{Q} - e_1)$, given a compensation plan $(\mathcal{Q}, \mathcal{B})$. This is because the self-control problem in period 2 is severe, and

---

[14] It is useful to note that this procrastination effect would also be observed with multiperiod quota plans even when the salesperson is an exponential discounter with $\beta = 1$. However, in this case, the firm will never benefit by offering a multiperiod quota plan and would offer single-period quotas. Thus, self-control issues are critical to the results in the paper. We discuss this in §4.4.

the salesperson in period 2 will not implement the plan because the compensation in period 2 does not motivate the salesperson to exert effort in the second period. The salesperson in period 1 therefore exerts more effort than $\psi(\beta m)$.

The second part of the proposition suggests that when the salesperson has a severe self-control problem, then the multiperiod contract can make both the firm and the salesperson better off. Intuition might suggest that in such cases, the firm will be in an even better position to exploit the salesperson's self-control problems. However, the results suggest that the opposite is true. Also note that this result is in contrast to the case when (10) does not bind, where the firm is able to exploit the salesperson's time inconsistency and reduces the salesperson's long-run surplus. This is because for low values of β, in a single-period plan, the salesperson exerts little effort, and the firm finds it too costly to induce effort. However, using a multiperiod quota plan, the firm finds it profitable to induce more effort, and the salesperson's compensation improves. Thus, for low values of β, multiperiod quotas can be a win–win for both the firm and the salesperson.

Revisiting our example when $h(e) = e^2/2$, consider the case when $\beta = 0.5$. In this situation, the salesperson exerts effort of $0.5m$ in each period with a single-period plan. In contrast, under the multiperiod quota plan, the salesperson exerts higher level of efforts: $0.603m$ in period 1 and $0.853m$ in period 2. Furthermore, if the firm uses a multiperiod quota plan, then its sales and profits increase by more than 45%. Also, the salesperson's long-run surplus improves by more than 45% under the multiperiod quota plan.

Propositions 2 and 3 show that the firm can benefit by delaying payments and paying a single bonus at the end. A natural question is whether the firm could do even better by specifying more complicated plans in which the firm could specify outputs for each period. This is not the case. To see this, suppose that the firm would like the salesperson to exert effort e_1^* in period 1 and e_2^* in period 2. The firm could specify a contract in which it only pays a bonus if the salesperson exerts e_1^* in period 1 and e_2^* in period 2. Let $e_1^* + e_2^* = \tilde{\mathcal{Q}}$ in this alternate plan, and let the bonus in this case be denoted by $\tilde{\mathcal{B}}$. Note that unlike the case so far, $\tilde{\mathcal{B}}$ need not only depend on the cumulative sales. The firm's profits in this case are $m\tilde{\mathcal{Q}} - \tilde{\mathcal{B}}$. However, we still need to satisfy the participation constraints for the salesperson ((9) and (10)). Suppose (10) binds under this proposed plan; i.e., $\beta\tilde{\mathcal{B}} = h(e_2^*)$. In this case, if the firm offers the same plan in which the payment is based on the cumulative sales and not on e_1^* and e_2^* separately, then the salesperson either will choose to implement (e_1^*, e_2^*) or will choose an alternative effort profile (e_1^u, e_2^u), which minimizes the effort

from the perspective of the first-period salesperson, and $e_1^u + e_2^u = \tilde{\mathcal{Q}}$, while still ensuring that the salesperson in the second period will implement the plan. In the first case, the profits under the simpler plan are the same. In the second case, by the optimality of the salesperson's choice, it follows that

$$h(e_1^*) + \beta h(e_2^*) > h(e_1^u) + \beta h(e_2^u), \qquad (13)$$

but then the firm can improve profits over the $(\tilde{\mathcal{Q}}, \tilde{\mathcal{B}})$ plan by slightly reducing the bonus payment, which contradicts the purported optimality of $(\tilde{\mathcal{Q}}, \tilde{\mathcal{B}})$. Therefore, if (10) binds, then the simpler multiperiod plan we have examined is still optimal. If (10) does not bind, then similar logic shows that a plan, which only looks at cumulative sales, is optimal.

One feature of our results is that the salesperson is not paid each period. Can the firm get the same results while still paying the salesperson for effort in each period? In other words, how much of the compensation needs to be back-loaded to still achieve the same results? To address this, denote the payment that the salesperson receives for achieving a per-period quota by $\mathcal{S}$ and the additional bonus received for achieving a two-period quota by $\mathcal{B}$. We now examine what the minimum level of $\mathcal{B}$ that the firm needs to set should be so as to achieve the same results as in Propositions 2 and 3.

PROPOSITION 4. *If (10) does not bind, then the firm can receive the same profits as in Proposition 2 by paying the salespersons $\mathcal{S}$ every period if the salesperson achieves a sales of $\psi(\beta m)$ and a bonus of $\mathcal{B}$ if the salesperson achieves $\mathcal{Q} = \psi(m) + \psi(\beta m)$ at the end of the two periods where*

$$\mathcal{S} = \frac{h(\psi(\beta m)) - (1 - \beta)h(\psi(m))}{\beta}, \qquad (14)$$

$$\mathcal{B} = \frac{(2 - \beta)h(\psi(m)) - h(\psi(\beta m))}{\beta}; \qquad (15)$$

$\mathcal{B}$ *is decreasing in β. However, if (10) binds, then it is optimal to pay the salesperson only in period 3.*

This compensation plan is shown in Figure 2. Intuitively, as β increases, the amount that the salesperson is paid in each period increases, and less payment needs to be shifted toward the third period. This is consistent with our results so far. Indeed, without problems of self-control (i.e., when $\beta = 1$), the firm does not benefit by delaying payments. Figure 3 shows that the proportion of bonus required as a percentage of total compensation increases as β decreases, for the specific case when $h(e) = e^2/2$. For example, if $\beta = 0.7$, then the firm pays 68.06% of total compensation using a cumulative bonus. On the other hand, if $\beta = 0.9$, then $\mathcal{B}$ need only be 16.96% of the total compensation.

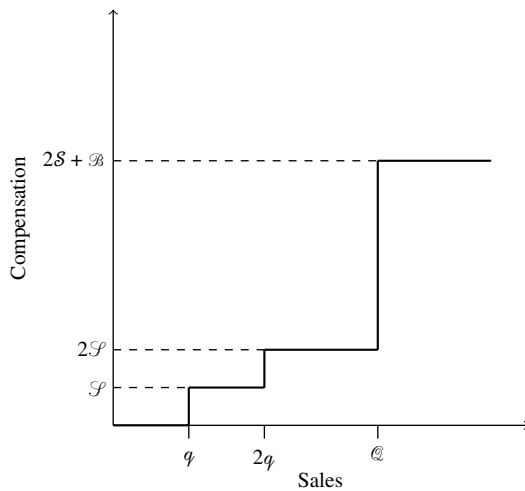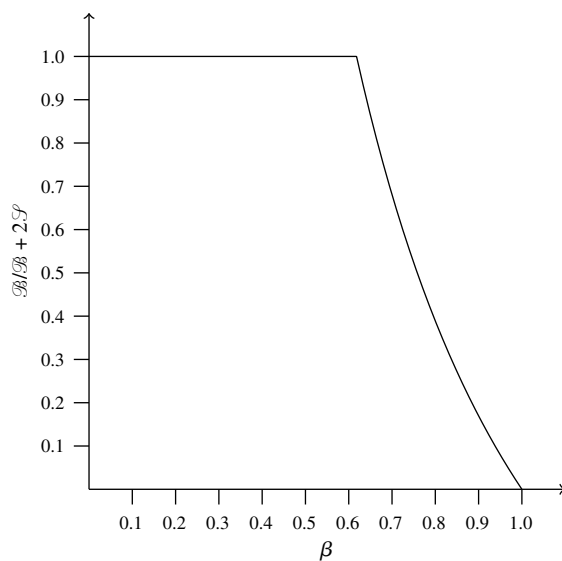**Figure 2   Alternate Compensation Plan When Equation (10) Does Not Bind**

Figure 3   Proportion of Total Compensation That Is Based on a Multiperiod Bonus

Note. $h(x) = x^2/2$.

4.   Model Extensions

In this section, we will relax some of the assumptions in the base model. This will allow us to see whether the results in the paper are robust to alternate assumptions and also allow us to generate additional insights. In §4.1, we relax the assumption that salespersons form rational expectations. This allows for the possibility that salespersons are not fully aware of their self-control problems. In §§4.2 and 4.3, we consider the cases when the firm cannot infer effort using the output. In §4.2, we analyze the case when salespersons are heterogeneous in their costs and the firm does not observe these costs. Salespersons could have different costs due to heterogeneity in their abilities. Alternatively, a salesperson may face different costs

due to different market conditions. The salesperson knows these costs but the firm only knows the distribution of these costs. Our analysis applies in either case. In such situations, because the firm does not observe the salesperson's costs, it is not able to infer the amount of effort exerted by examining the output.[15] In §4.3, we consider the case when output is only stochastically related to effort. In other words, the salesperson's efforts sometimes lead to sales and sometimes it does not, despite the salesperson exerting effort. Again, in this case, the firm cannot infer effort from the output. Finally, in §4.4, we consider the case when the salesperson uses both hyperbolic and exponential discounting.

4.1.   Salespeople Do Not Form Rational Expectations

In the base model, we assumed that salespersons fully realize their self-control problems and can accurately predict what they will do in subsequent periods. Casual evidence, however, suggests that individuals are often overoptimistic. To model this, we can assume that the salesperson at time 1 believes that his future self in time period 2 will have a hyperbolic discount parameter $\tilde{\beta} \geq \beta$. This approach of modeling incorrect expectations has been used by numerous other authors (see, for example, O'Donoghue and Rabin 2001, DellaVigna and Malmendier 2004). If $\tilde{\beta} = \beta$, then the salesperson has rational expectations. If $\tilde{\beta} = 1 > \beta$, then the salesperson at time period 1 incorrectly believes that he does not have any self-control problem in period 2.

First, consider the case when (10) does not bind. In this case, rational expectations play no role in the analysis, and Proposition 2 continues to hold. If $h(\cdot)$ is quadratic, then we know from Lemma 1 that the results would hold as long as $\beta > 0.618$. If (10) does bind, then the analysis would change. In particular, in such situations, the salesperson in period 1 thinks that the maximum effort that the second-period salesperson is willing to exert is given by

$$e_2 = h^{-1}(\tilde{\beta}\mathscr{B}) > h^{-1}(\beta\mathscr{B}). \tag{16}$$

In such a situation, the firm can exploit the salesperson's inability to rationally anticipate his action. In particular, consider the case when the firm chooses

[15] In such turbulent environments, multiperiod quota plans can provide another benefit, which we do not consider in this paper. If sales are stochastic, then the total sales over multiple time periods could give a better estimate of the salesperson's productivity and effort. This can then be used to weed out ineffective salespersons. Fernández-Gaucherand et al. (1995) develop a model in which the firm uses sales quotas to determine a salesperson's productivity. However, in their model, quotas are only used for termination decisions and are not related to compensation. Furthermore, in their model, salespersons are not strategic.

a pair $(\mathcal{Q}, \mathcal{B})$ such that there exist $e_1(\mathcal{Q})$ and $e_2(\mathcal{Q})$ such that $\mathcal{Q} = e_1(\mathcal{Q}) + e_2(\mathcal{Q})$ and

$$\beta\mathcal{B} \geq h(e_1(\mathcal{Q})) + \beta h(e_2(\mathcal{Q})), \qquad (17)$$

$$\tilde{\beta}\mathcal{B} > h(e_2(\mathcal{Q})) > \beta\mathcal{B}, \qquad (18)$$

where

$$h'(e_1(\mathcal{Q})) = \beta h'(\mathcal{Q} - e_1(\mathcal{Q})) = \beta h'(e_2(\mathcal{Q})). \qquad (19)$$

Equations (17) and (18) ensure that the salesperson will be willing to exert the effort in period 1, believing that in period 2, he will make the required effort to make the quota. Given (18), the salesperson chooses effort in the first period using (19). Note that the firm can always choose a $\mathcal{B}$ such that the two inequalities in (18) are satisfied. Also, it is easy to see that given (18), the salesperson believes that $(e_1(\mathcal{Q}), e_2(\mathcal{Q}))$ is implementable, when, in fact, it is not. This implies that the firm can make the salesperson exert effort in period 1, without paying the bonus amount. Since $e_1(\mathcal{Q})$ is strictly increasing in $\mathcal{Q}$, it follows that the firm can exploit naïve salespeople by making them exert an arbitrary effort in the first period and no effort in the second period. Note, however, that these exploitative contracts are only feasible when the value of β is relatively low.[16]

4.2. Heterogeneous Salesforce
Now consider the case when there are two types of salespersons with different costs of efforts. The cost function is given by $\mu_i h(\cdot)$, where μ_i is a constant. We normalize $\mu_1 = 1$. The second segment faces higher cost function $\mu_2 h(\cdot)$, where $\mu_2 > 1$. The proportion of the first type of salesperson in the market is $0 \leq \alpha \leq 1$. The firm does not observe the salesperson's type but can potentially offer a menu of contracts so that salespersons can self-select. Note that the formulation can also be used to represent the case where the salesperson knows more about the market conditions than the firm. With probability α, the market is favorable, and the salesperson incurs lower costs, while with probability $(1 - \alpha)$, the market is tougher and requires a higher level of effort. At time 0, the salesperson observes the market condition, but the firm only knows the prior probabilities.

First, consider the case of a single-period plan. The usual solution to this case is that the salespersons are offered a menu of quota plans and they self-select which one they will implement. Thus, at time 0, the

firm offers a quota q_i for the salesperson of type i with an associated bonus of ℓ_i. If the salesperson i exerts effort q_i, then the plan must satisfy the individual rationality constraint and the incentive compatibility constraints. In other words, the salesperson must be willing to exert the desired effort, and the salesperson of type i should be willing to implement q_i and not q_j, $j \neq i$.[17] The relevant constraints then are

$$\beta\ell_i \geq \mu_i h(q_i), \quad i = 1, 2, \qquad (20)$$

$$\beta\ell_i - \mu_i h(q_i) \geq \beta\ell_j - \mu_i h(q_j), \quad i = 1, \dots, 2, j = 3 - i, \quad (21)$$

where (20) are the individual rationality constraints and (21) are the incentive compatibility constraints for period 1.[18] In this case, the individual rationality constraint for the less efficient salesperson, i.e., salesperson 2 binds, and the incentive compatibility constraint for salesperson 1 in period 1 binds. This leads to

$$\ell_2 = \frac{\mu_2 h(q_2)}{\beta}, \qquad (22)$$

$$\ell_1 = \ell_2 - \frac{h(q_2) - h(q_1)}{\beta} = \frac{(\mu_2 - 1)h(q_2) + h(q_1)}{\beta}. \qquad (23)$$

The problem for the firm then is to choose the quotas such that

$$\{q_1, q_2\} \in \operatorname*{arg\,max}_{q_1, q_2} \left\{ \alpha\left[mq_1 - \frac{(\mu_2 - 1)h(q_2) + h(q_1)}{\beta} \right] \right.$$
$$\left. + (1 - \alpha)\left[mq_2 - \frac{\mu_2 h(q_2)}{\beta} \right] \right\}. \qquad (24)$$

It is easy to see then that $q_1 = \psi(m\beta)$ and $q_2 = h((m\beta(1 - \alpha))/(\mu_2 - \alpha))$. Comparing with the case with no heterogeneity, we see that the quota for the more efficient salesperson remains the same while that of the less efficient salesperson is reduced.[19] This result is consistent with earlier work using time-consistent preferences (e.g., Rao 1990). The intuition is that the presence of less efficient salespersons prevents the firm from extracting all the surplus from the more efficient salesperson, who can always get positive

---

[16] Of course, in the long run, salespeople are unlikely to be fooled by such contracts. Firms may also not want to implement such exploitative contracts because of concerns of fairness, reputation, employee satisfaction, and turnover (see also O'Donoghue and Rabin 1999a for a discussion of these issues). These aspects are, however, beyond the scope of the current paper.

[17] Note that we assume that the firm does not force the salesperson to choose a quota ex ante but only publishes the menu of contracts, and the salesperson is paid according to the sales that he achieves. Alternatively, the firm could force the salesperson to commit to a contract and not pay anything even if he achieves the lower quota. Note that this distinction is irrelevant in models of time-consistent preferences, but it does matter in our context. We make the conservative assumption that the firm does not force the salesperson to ex ante choose a contract because it reflects current practice.

[18] As before, it is easy to see that the incentive compatibility and individual rationality constraints for period 1 are stronger than for period 0, and therefore we need only consider period 1 constraints.

[19] To see this, note that this salesperson would be assigned a quota of $h((m\beta)/\mu_2)$ absent heterogeneity and note the fact that $(1 - \alpha)/(\mu_2 - \alpha) < 1/\mu_2$.

surplus by choosing the quota designed for the less efficient salesperson. To reduce the rent to the more efficient salesperson, the firm reduces the quota and the payment to the less efficient salesperson.

Now, consider the multiperiod plan. First, note that the logic of Proposition 1 still holds, and therefore multiperiod plans continue to dominate in this case. We now characterize the plan. In this situation, the firm can offer two types of menus $(\mathbb{Q}_1, \mathscr{B}_1)$ and $(\mathbb{Q}_2, \mathscr{B}_2)$, so that segment 1 chooses the menu $(\mathbb{Q}_1, \mathscr{B}_1)$, and segment 2 chooses the other menu (see Figure 4). For this to work, we need the individual rationality constraints and incentive compatibility conditions to be satisfied for each of the two segments. Denote the effort functions for salesperson i in period j, given a plan $(\mathbb{Q}_i, \mathscr{B}_i)$, by $e_{ji}(\mathbb{Q}_i)$. Define

$$\zeta_i(\mathbb{Q}) \equiv \mu_i \left[\frac{h(e_{1i}(\mathbb{Q}))}{\beta} + h(\mathbb{Q} - e_{1i}(\mathbb{Q})) \right]. \quad (25)$$

Thus, $\zeta_i(\mathbb{Q})$ is the effort cost for the salesperson i at time 1 if he implements quota $\mathbb{Q}$. The individual rationality constraints for periods 1 and 2 are

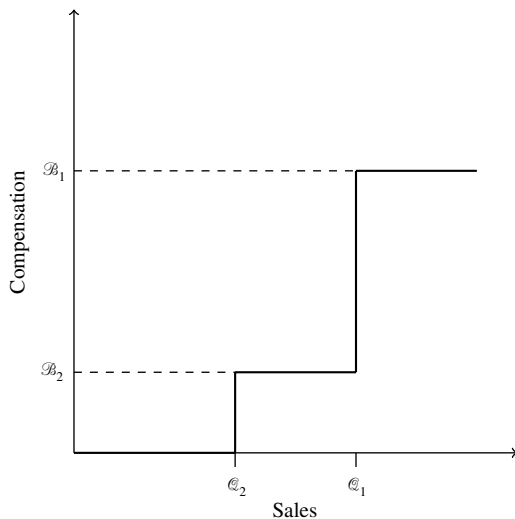$$\mathscr{B}_i \geq \zeta_i(\mathbb{Q}_i), \quad i = 1, 2, \quad (26)$$

$$\mathscr{B}_i \geq \frac{\mu_i h(\mathbb{Q}_i - e_{1i}(\mathbb{Q}_i))}{\beta}, \quad i = 1, 2. \quad (27)$$

The corresponding incentive compatibility conditions for periods 1 and 2 are

$$\mathscr{B}_i - \zeta_i(\mathbb{Q}_i) \geq \mathscr{B}_j - \zeta_i(\mathbb{Q}_j), \quad i = 1, \ldots, 2, \ j = 3 - i, \quad (28)$$

$$\mathscr{B}_i - \frac{\mu_i h(\mathbb{Q}_i - e_{1i}(\mathbb{Q}_i))}{\beta} \geq \mathscr{B}_j - \frac{\mu_i h(\mathbb{Q}_j - e_{1i}(\mathbb{Q}_i))}{\beta}, \quad i = 1, \ldots, 2, \ j = 3 - i. \quad (29)$$

Figure 4　Compensation Plan When Salespersons Are Heterogeneous

We consider the case when both salespersons are active in the market. In such a case, clearly (26) cannot bind for salesperson 1 because salesperson 1 can always get positive surplus by switching to plan $(\mathbb{Q}_2, \mathscr{B}_2)$. Also, the incentive compatibility condition for salesperson 2 does not bind. First, consider the case when the second-period individual rationality and incentive compatibility constraints do not bind for both segments. If $h(e) = e^2/2$, then this is true if $\beta < 0.61$ and $\mu_2 > (1 - 2\alpha(1 - \beta))/(1 - 2\beta)$. We have the following result.

Proposition 5. *If (27) and (29) do not bind, then*
a. $\mathbb{Q}_1^* = \psi(m) + \psi(m\beta) > \mathbb{Q}_2^* = \psi(((1 - \alpha)m)/(\mu_2 - \alpha)) + \psi((\beta(1 - \alpha)m)/(\mu_2 - \alpha))$.
b. *If $h(e) = e^2/2$, then $\mathscr{B}_1^*/\mathbb{Q}_1^* > \mathscr{B}_2^*/\mathbb{Q}_2^*$. Furthermore, as β increases, there is more distortion of quota for the high-cost segment.*

The optimal plan is therefore of the type shown in Figure 4. Proposition 5 shows that the presence of a segment of salespersons who have higher costs leads to lower goals being assigned to the salesperson with higher effort costs. Using our results in Proposition 2, we see that the optimal quota for the more efficient salesperson remains unchanged, whereas the quota for the less efficient salesperson is reduced. Also, notice that consistent with our earlier results, using the multiperiod quota plans, the firm is able to induce optimal effort from salesperson 1 in period 2. Furthermore, as before, the total effort induced using the multiperiod quota plans exceeds that by using single-period quota plans. The next part of Proposition 5 shows that the firm pays the more efficient salesperson at a much higher rate than the less efficient salesperson. For example, consider the case when $h(e) = e^2/2$. If $\beta = 0.7$, $\mu_2 = 2$, $m = 4$, and $\alpha = 0.5$, the firm pays 3.02 for achieving a quota of 2.26 but pays at a much higher rate, i.e., 15.11 for achieving a quota of 6.8. This mirrors many quota plans identified in empirical research.

The last part of Proposition 5 shows how β affects the degree of distortion in the quota plan offered to the less efficient group. We find that if $h(\cdot)$ is quadratic, then as the self-control problem increases (i.e., β decreases), the need to distort the quota for the less efficient salesperson decreases. In other words, self-control problems mitigate quota distortion due to salesforce heterogeneity.

Now, consider the case when the second-period constraints bind. If only the second-period individual rationality constraints bind, then as before, the firm will only distort the quota for the less efficient salesperson. However, if the incentive compatibility constraint for salesperson 1 binds in period 2, then unlike standard models, the firm will have to distort the quotas for both the salespersons. Nevertheless, even in

this case, multiperiod quota plans will be more profitable than single-period quota plans.

4.3. Sales Are Stochastically Related to Effort

Now consider the case when the salesperson's effort is only stochastically related to sales. Consequently, the firm cannot infer effort from observed sales. One approach to model this is to assume that the salesperson's efforts do not always lead to sales and the probability that effort leads to a successful outcome is increasing in effort. We assume that the salesperson can choose between two levels of effort (e_h, e_l), where $e_h > e_l$. The lower level of effort can be viewed as the base level of effort (such as routine calling of clients) that the firm can observe and compensate with a salary. However, the firm cannot observe whether the salesperson has made additional effort, which increases the chances of success. The cost of this additional effort is $c_h > c_l$. We normalize $c_l = 0$ and set $c_h = c > 0$. To model the idea that effort increases the probability of success, we assume that a low level of effort, i.e., e_l, leads to success (output of 1) with probability p_l. However, the success probability with a high level of effort is p_h, where $p_h > p_l$. Thus, although the firm cannot observe effort, if sales do occur, then it is more likely that the salesperson has exerted a high level of effort. We will assume that both the firm and the salesperson are risk neutral. This considerably simplifies the analysis and allows us to focus on the role of uncertainty and present-biased preferences.[20]

First, consider the firm's problem when it offers a compensation plan based on each period's output. The firm can motivate the salesperson to exert high level of effort if the bonus ℓ_1 is such that

$$\beta \ell_1 p_h - c \geq \beta \ell_1 p_l. \tag{30}$$

In other words, we need

$$\ell_1 \geq \frac{c}{\beta(p_h - p_l)}. \tag{31}$$

If the firm wants the salesperson to exert a high level of effort, then the firm will set the quota to be one unit in each period and set ℓ_1 such that (31) is satisfied with equality. However, for the firm to want to induce high effort, we need

$$m p_h - \frac{c}{\beta(p_h - p_l)} \geq m p_l, \tag{32}$$

which implies that we need

$$m \geq \frac{cp_h}{\beta(p_h - p_l)^2} = m_0. \tag{33}$$

Now, consider the case when the firm sets a multiperiod quota. We have the following result.

PROPOSITION 6. *The firm strictly benefits by offering a multiperiod quota plan with $\mathcal{Q} = 2$ and $\mathcal{B} = (c(1 + \beta(p_h - p_l)))/(\beta p_h (p_h - p_l))$ if $m \in (m_1, m^*)$, where $m^* = (m_0(1 - \beta(p_h - p_l)))/(1 - p_h)$ and $m_1 = (m_0(1 + \beta(p_h - p_l)))/(1 + p_h)$. Furthermore, the range of m over which the multiperiod plan dominates the single-period plan increases in p_l and c, and it decreases in β.*

To understand Proposition 6, first note that when $p_h = 1$ and $p_l = 0$, we have the base case. In this case, the firm cannot benefit by offering a multiperiod quota plan unless $\beta < 1$. Furthermore, in this case, the firm *always* benefits by offering a multiperiod quota plan, as long as $\beta < 1$. This is consistent with Proposition 1. The intuition, as before, is that the firm can pay relatively less to the salesperson by exploiting his time-inconsistent preferences. With uncertainty, additional forces are in play that we need to understand. First, note that under the multiperiod contract, the salesperson will not exert effort in the second period if there is no sales in period 1 (see Chung et al. 2008 for empirical evidence of this phenomenon). This aspect hurts the firm's revenues, and the effect is stronger as m increases. Thus, for high-margin items, multiperiod quotas are less beneficial when sales are stochastic.

Now, consider the impact of p_l. As p_l increases, the chances that the salesperson could get sales without high effort increases. This increases the attractiveness of the low-effort option. In the case of single-period contracts, the firm needs to compensate the salesperson for the cost of high effort in every period relative to the cost of low effort. However, for multiperiod contracts, with a quota of 2, the salesperson exerts high effort in the second period only if sales occur in the first period. Thus, the expected cost for the salesperson from the first-period perspective is lower when he chooses to exert high effort in period 1. This implies that the firm needs to offer a relatively lower bonus as p_l increases when it uses a two-period contract as opposed to a one-period contract. Furthermore, as p_l increases, the chance that sales will occur without high effort in period 2 also increases. This mitigates the problem with multiperiod contracts, that they can sometimes lead to no effort in period 2. Consequently, as p_l increases, a multiperiod contract becomes more attractive.

As c increases, the cost of inducing effort from the salesperson increases, and therefore the firm must pay a higher bonus in order to induce effort. Under a multiperiod plan, the firm, however, pays the bonus only

---

[20] Risk aversion is likely to make multiperiod contracts less attractive because it increases the chance that the salesperson will not be paid at the end of two periods. It is well known in the literature that risk aversion leads to a fixed salary component in the salesforce compensation plan (see, for example, Basu et al. 1985). Nevertheless, the basic intuition in our analysis would still hold in this more complicated case.

if the salesperson is able to sell in both periods. This aspect favors the multiperiod quota plan, and as c increases, the firm pays out relatively less under the multiperiod quota plan. The effect of β is consistent with our earlier results. An increase in β makes the time-inconsistency problem less severe, and therefore multiperiod contracts become less attractive.

A natural question is whether the firm can do better by offering compensation when there is no sales in period 1. In other words, the firm could offer compensation ℓ for sale of one unit and $\mathscr{B}$ for sale of two units, where $\mathscr{B}$ does not need equal 2ℓ. However, it is easy to show that this is not the case, and whenever multiperiod contracts are beneficial, they continue to dominate single-period contracts. Now, let us consider the impact of multiperiod contracts on salesperson's welfare. We have the following result.

PROPOSITION 7. *The salesperson's long-run surplus increases with the multiperiod contract when $m \in (m_1, m_0)$ and decreases if $m > m_0$.*

Proposition 7 shows that as before, multiperiod contract plans can sometimes be a win–win for all parties. When $m \in (m_1, m_0)$, the firm is able to motivate the salesperson to exert a high level of effort under the multiperiod contract. However, under a single-period contract, the firm does not find it profitable to induce a high level of effort when $m \in (m_1, m_0)$. Note that as β increases, (m_1, m_0) decreases. Thus, the region for which salesperson's long-run surplus improves due to multiperiod contracts is larger for lower values of β, i.e., when the self-control problem is more severe. This is intuitive because single-period contracts are less likely to be able to induce effort when β is low. Finally, note that when m is sufficiently high, salespersons are worse off under the multiperiod plan. This is because in this situation, the firm in any case is willing to pay in order to induce the salesperson to exert high effort. However, the two-period contract enables the firm to exploit the salesperson's time-inconsistent preferences and pay them less. Furthermore, the presence of uncertainty enables the firm to pay the salesperson only when sales occur in both periods. Both of these effects reduce the salesperson's long-run surplus.

4.4. Salesperson's Exponential Discount Factor Is Less Than 1

In the base model, we assumed that the per-period exponential discount factor is 1. Now let us consider the more general case when $\delta < 1$. In other words,

$$D(\tau) = \begin{cases} 1 & \text{if } \tau = 0, \\ \beta\delta^\tau & \text{otherwise.} \end{cases} \qquad (34)$$

For simplicity, we assume that the firm does not discount future profits. This allows us to study the role

of two types of discounting by the salesperson, i.e., exponential discounting and the hyperbolic discounting. First, consider the case when $\delta < 1$ and $\beta = 1$. The optimal single-period contract leads to an effort of $\psi(m\delta)$ in each period. However, under the optimal multiperiod quota plan, the salesperson exerts $\psi(m\delta^2)$ in period 1 and $\psi(m\delta)$ in period 2 (see the technical appendix for the proof). Note that with a multiperiod quota plan, salespersons exert more effort in the second period compared with period 1. Thus, the procrastination effect we observed with β can be observed even when salespersons only use exponential discounting. However, when $\beta = 1$, procrastination by salespersons unambiguously hurts the firm's profits, and the firm would strictly prefer to pay for output after every period (see the technical appendix for the proof). Recall, however, that when $\beta < 1$, the salesperson's time-inconsistent preferences could make it more attractive for the firm to defer payment. Thus, exponential discounting and hyperbolic discounting affect the optimal contract in opposite ways.

Now consider the case when both δ and β are strictly less than 1. Let us first examine the simple case when the firm can either offer a contract in which the payment is only made in period 3 or use single-period contracts. In this case, we show in the technical appendix that a sufficient (but not necessary) condition for using the pure multiperiod contract is that $\delta > 1/(2 - \beta)$. For example, if $\beta = 0.8$, then the pure multiperiod contract dominates as long as $\delta > 0.83$, which is not too strong a condition. However, multiperiod quotas can be useful even when this condition is not satisfied. We have the following proposition.

PROPOSITION 8. *If $\beta < 1$ and $\delta < 1$, then there always exists a multiperiod contract such as in Figure 2, which strictly dominates any single-period contract.*

The results therefore show that when the salesperson uses both types of discounting, the firm uses a mixture of per-period payment and bonus payments at the end. In other words, even when $\delta < 1$, as long as $\beta < 1$, the firm always benefits by back-loading at least some compensation in the form of bonus payments. Thus, the basic intuition from our earlier analysis continues to hold in the case of $\delta < 1$.

5. Conclusion

The purpose of this paper was to examine an optimal compensation plan when salespersons have self-control problems. We model salespersons' self-control problems by assuming that they have present-biased preferences. We then examine how these preferences impact compensation plans. Our results suggest that in the presence of present-biased preferences, firms can sometimes increase their profits by delaying payments using multiperiod quota plans. We find that

such plans can lead to effort distortions such as procrastination or reduced effort in later periods. Despite these negative consequences of multiperiod quotas, we find that such plans can often motivate salespersons to exert more effort and improve firms' profits. Our analysis reveals that when salespersons have low levels of self-control problems, multiperiod quota plans will decrease their long-run surplus. This is because firms use the multiperiod quota plans to make the salesperson work harder and do not have to pay him as much for the increased effort. This happens because multiperiod quotas enable the firm to exploit the time inconsistency in the salesperson's present-biased behavior. In contrast, when salespersons have high levels of self-control problems, multiperiod quota plans can improve their long-run surplus. This is because absent such plans, the salesperson will exert very little effort, and the firm will not find it beneficial to elicit more effort. Consequently, in such situations, the salesperson will receive low levels of compensation in single-period plans. Under multiperiod quota plans, however, the firm finds it profitable to induce more effort, and the salesperson also gets paid more. Our paper therefore establishes a motivational role for multiperiod quota plans and clarifies why delaying payments by using annual bonuses may be optimal.

Future research can extend our research in multiple ways. Our results suggest that firms can benefit by delaying at least some payments in the form of bonuses. An important question is, how can firms structure such bonuses? And what is the optimal length of bonus periods? In practice, firms use quarterly bonuses as well as annual bonuses. Absent exponential discounting, our results suggest that firms will always delay payments to the last period. However, when we consider exponential discounting by a salesperson, our results in §4.4 show that such a strategy may not always be optimal, and firms may prefer to pay at least some portion of the payment each period. These results suggest that in deciding the optimal length of time, the firm must balance the opposing effects of the two types of discounting by the salesperson, i.e., the exponential discounting and hyperbolic discounting. Whereas hyperbolic discounting works toward extending the length of time after which bonuses are paid, exponential discounting would tend to shorten it.[21]

We have considered a situation in which the salesperson makes only one type of decision, i.e., how much effort to exert to achieve sales. In a more general case, we could have a situation in which the salesperson needs to not only decide how much effort to exert but also choose where such efforts should be directed. More generally, we could extend our model to consider principal–agent formulation in which the agent chooses the timing and the quantity of effort to exert from a menu of tasks. It would be interesting to study how to design optimal compensation plans in such contexts. In our formulation, we have assumed that the salespersons' present-biased preferences do not change over time. However, one could envision the case that with experience, salespersons may learn to better self-manage their time. This would then imply that quota-bonus plans are more effective for less experienced salespersons, something that could be empirically examined. Our paper provides several other hypotheses that can be subjected to empirical tests. For example, our results suggest that bonuses are more likely to induce higher efforts from individuals with higher self-control problems. Our results also suggest that if effort costs are high, then multiperiod quotas are more profitable. Future research can empirically assess the validity of these results.

Electronic Companion
An electronic companion to this paper is available as part of the online version that can be found at http://mktsci.journal.informs.org/.

Acknowledgments

References
Akerlof GA (1991) Procrastination and obedience. *Amer. Econom. Rev.* 81(2):1–19.

Amaldoss W, Jain S (2005) Pricing of conspicuous goods: A competitive analysis of social effects. *J. Marketing Res.* 42(1):30–42.

Basu AK, Lal R, Srinivasan V, Staelin R (1985) Salesforce compensation plans: An agency theoretic perspective. *Marketing Sci.* 4(4):267–291.

Benzion U, Rapoport A, Yagil J (1989) Discount rates inferred from decisions: An experimental study. *Management Sci.* 35(3):270–284.

Bernheim B, Rangel A (2005) Behavioral public economics: Welfare and policy analysis with non-standard decision makers. NBER Working Paper 11518, National Bureau of Economic Research, Cambridge, MA.

---

[21] For example, consider the case when the salesperson makes effort decisions three times, and $h(e) = e^2/2$, $\beta = 0.9$, and $\delta = 0.9$. In this situation, using single-period quotas dominates using only a bonus payment in period 4. In particular, using an optimal quota of 0.81 per period and a per-period bonus of 0.42, the firm makes profits equal to 1.215. In contrast, if the firm uses a quota of 2.36 in period 4, it only makes profits of 1.18. However, offering only a bonus becomes attractive if β decreases to 0.8 or δ rises to 0.95.

Nevertheless, if we allow for a mixture of quota and per-period payments, then the optimal plan still involves per-period payments and an additional bonus based on cumulative sales, consistent with our results in Proposition 8. In particular, in this case, the firm can use different bonuses for periods 2–4 and increase its profits to 1.29.

Carpenter G, Nakamoto K (1990) Competitive strategies for late entry into a market with a "dominant brand." *Management Sci.* 36(10):1268–1279.

Chapman G (1996) Temporal discounting and utility for health and money. *J. Experiment. Psych.: Learning, Memory, Cognition* 22(3):771–791.

Chung D, Steenburgh T, Sudhir K (2008) Do bonuses enhance sales productivity? A dynamic structural analysis of bonus-based compensation plans. Working paper, Yale School of Management, New Haven, CT.

Dai T, Jerath K (2011) Salesforce compensation with inventory considerations. Working paper, Carnegie Mellon University, Pittsburgh.

DellaVigna S, Malmendier U (2004) Contract design and self-control: Theory and evidence. *Quart. J. Econom.* 129(2):353–402.

Fernández-Gaucherand E, Jain S, Lee HL, Rao AG, Rao MR (1995) Improving productivity by periodic performance evaluation: A Bayesian stochastic model. *Management Sci.* 41(10):1669–1678.

Frayne CA, Geringer JM (2000) Self-management training for improving job performance: A field experiment involving salespeople. *J. Appl. Psych.* 85(3):361–372.

Fudenberg D, Levine D (2006) A dual-self model of impulse control. *Amer. Econom. Rev.* 96(5):1449–1476.

Gilpatric S (2008) Present-biased preferences, self-awareness and shirking. *J. Econom. Behav. Organ.* 67(3–4):735–754.

Gruber J, Köszegi B (2001) Is addiction rational? Theory and evidence. *Quart. J. Econom.* 116(4):1261–1305.

Gul F, Pesendorfer W (2001) Temptation and self control. *Econometrica* 69(6):1403–1435.

Harris C, Laibson D (2002) Hyperbolic discounting and consumption. Dewatripont M, Hansen LP, Turnovsky S, eds. *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, Vol. 1 (Cambridge University Press, Cambridge, UK), 258–298.

Jain S (2009) Self-control and optimal goals: A theoretical analysis. *Marketing Sci.* 28(6):1027–1045.

Jain S (2012) Marketing of vice goods: A strategic analysis of the package size decision. *Marketing Sci.* 31(1):36–51.

Joseph K, Kalwani MU (1998) The role of bonus pay in salesforce compensation plans. *Indust. Marketing Management* 27(2):147–159.

Kirby KN (1997) Bidding on the future: Evidence against normative discounting of delayed rewards. *J. Experiment. Psych.: General* 126(1):54–70.

Laibson D (1997) Golden eggs and hyperbolic discounting. *Quart. J. Econom.* 112(2):443–477.

Lal R, Srinivasan V (1993) Compensation plans for single- and multi-product salesforces: An application of the Holmstrom-Milgrom model. *Management Sci.* 39(7):777–793.

Machado F, Sinha R (2007) Smoking cessation: A model of planned vs. actual behavior for time-inconsistent consumers. *Marketing Sci.* 26(6):834–850.

Misra S, Nair HS (2011) A structural model of sales-force compensation dynamics: Estimation and field implementation. *Quant. Marketing Econom.* 9(3):211–257.

O'Donoghue T, Rabin M (1999a) Doing it now or doing it later. *Amer. Econom. Rev.* 89(1):103–124.

O'Donoghue T, Rabin M (1999b) Incentives for procrastinators. *Quart. J. Econom.* 114(3):769–816.

O'Donoghue T, Rabin M (2000) Economics of immediate gratification. *J. Behav. Decision Making* 13(2):233–250.

O'Donoghue T, Rabin M (2001) Choice and procrastination. *Quart. J. Econom.* 116(1):121–160.

O'Donoghue T, Rabin M (2003) Self-awareness and self-control. Loewenstein G, Read D, Baumeister R, eds. *Time and Decision: Economic Perspectives on Intertemporal Choice* (Russell Sage Foundation, New York), 217–243.

Oyer P (1998) Fiscal year ends and nonlinear incentive contracts: The effect on business seasonality. *Quart. J. Econom.* 113(1):149–185.

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

Raju JS, Srinivasan V (1996) Quota-based compensation plans for multiterritory heterogeneous salesforces. *Management Sci.* 42(10):1454–1462.

Rao RC (1990) Compensating heterogeneous salesforces: Some explicit solutions. *Marketing Sci.* 9(4):319–341.

Renn R, Allen D, Huning T (2011) Empirical examination of the individual-level personality-based theory of self-management failure. *J. Organ. Behav.* 32(1):25–43.

Steenburgh T (2008) Effort or timing: The effect of lump-sum bonuses. *Quant. Marketing Econom.* 6(3):235–256.

Syam N, Krishnamurthy P, Hess JD (2008) That's what I thought I wanted? Miswanting and regret for a standard good in a mass-customized world. *Marketing Sci.* 27(3):379–397.

Thaler RH (1981) Some empirical evidence on dynamic inconsistency. *Econom. Lett.* 8(3):201–207.

Thaler RH, Shefrin HM (1981) An economic theory of self-control. *J. Political Econom.* 89(2):392–406.

Villas-Boas JM (2009) Product variety and endogenous pricing with evaluation costs. *Management Sci.* 55(8):1338–1346.

Wernerfelt B (1995) A rational reconstruction of the compromise effect: Using market data to infer utilities. *J. Consumer Res.* 21(4):627–633.

Wu Y, Ramachandran K, Krishnan V (2009) Managing projects with present-biased agents. Working paper, National University of Singapore, Singapore.