

Optimal Sales Force Compensation in Dynamic Settings: Commissions versus Bonuses*

Anja Schöttner[†]

Abstract

This paper studies optimal sales force compensation plans in a multi-period moral-hazard model when the firm wants to implement high effort in every period but only obtains aggregate information on sales. The sales agent chooses effort each period after observing previous sales and his incentive responsiveness might change over time. The paper derives conditions under which a linear incentive scheme - a pure commission - dominates a bonus plan and vice versa. A commission is optimal if the agent is most difficult to motivate in the last period. Otherwise, combining the commission with a bonus plan can lower the firm's cost of providing incentives in earlier periods. The results are robust to different types of cost externalities and demand externalities across periods. However, if the firm obtains intermediate sales information, bonus plans dominate commissions.

Key Words: sales force compensation, linear incentive contracts, commissions, quota-based bonuses

*I would like to thank Matthias Kräkel for very helpful discussions and comments.

[†]Humboldt University Berlin, School of Business and Economics, Spandauer Str. 1, D-10099 Berlin, Germany, tel: +49 30 20935716, e-mail: anja.schoettner@hu-berlin.de.

1 Introduction

Personal selling via sales forces is one of the most important marketing instruments. According to Zoltners et al. (2008, p. 115), U.S. firms spend approximately \$800 billion on sales forces each year – almost three times as much as they spent on advertising in 2006. Sales force compensation plans, however, differ across firms. Joseph and Kalwani (1998, p. 149) report that 5% of the 266 companies participating in a survey exclusively pay fixed salaries to their salespeople, 24% use only commissions, 37% use only a bonus component, and 35% use both commissions and bonus pay. By far the most important criterion in determining bonus payments was the comparison of actual sales and a predetermined quota. Commissions and/or quota-based bonuses thus appear to be the most common forms of sales force compensation. Commissions are linear incentive schemes that reward each sale equally, whereas bonuses as non-linear compensation forms emphasize the importance of reaching specific goals. Empirical studies suggest that these different types of compensation plans indeed have different impacts on sales force motivation and productivity. Kishore et al. (2013) analyze data from a pharmaceutical corporation that switched from a bonus plan to a commission plan, thereby increasing overall productivity by 24%. By contrast, Steenburgh (2008) shows that quota bonuses can effectively provide incentives despite a possible discouragement effect, which arises when salespeople learn that they are unlikely to make quota. Chung et al. (2014) find that quota-based bonuses enhance performance in a firm that also uses commissions.

These results suggest that best practices for sales force compensation depend on the specific characteristics of the firm and its environment. It is therefore important to understand what drives the optimality of linear forms of compensation, and when they are dominated by non-linear incentive schemes. The present paper investigates this question within an agency-theoretic framework, considering important dynamic aspects of sales force motivation that have been neglected in the literature so far. I analyze a multi-period moral-hazard model with binary effort and stochastic outcomes, where a sales agent decides whether to work hard or not to sell a firm's product or service every time he talks to a customer. One important aspect of the model is that the firm wants the agent to exert high effort with every customer but can only observe total sales at the end of the last period because it cannot monitor every customer contact. The agent, however, immediately learns whether or not he could sell the product and hence can adjust his effort to the sales history. Another important feature of the model is that the agent's incentive responsiveness varies over time due to changing effort costs or market characteristics. The agent is further protected by limited liability¹ and, as a consequence, usually earns a rent.

¹This assumption is common in the contract theoretical literature. See, e.g., Sappington (1983), Innes (1990),

I find that the optimal compensation plan crucially depends on how the agent’s incentive responsiveness changes over time. In a wide range of settings, a pure commission scheme is optimal whenever the agent is hardest to motivate with the last customer. A bonus cannot motivate the agent towards the end of the season if it turns out that he cannot make quota anymore. Hence, a bonus is least effective when the agent is most difficult to incentivize. In contrast, if the agent is harder to motivate in an earlier period, the firm should combine the commission with a bonus plan. Bonuses can effectively motivate the agent in early periods because they focus rewards on high aggregate sales outcomes, which the agent can always attain at the start of the season. The difficulty of motivating the agent in a given period is measured by the cost-responsiveness ratio, i.e., the ratio between the agent’s effort cost increase when working hard and the sales-effort responsiveness. If this ratio is highest in the last period, then a pure commission is optimal. For instance, increasing effort costs and a decreasing sales-effort responsiveness favor a linear compensation scheme. Sales might become less responsive to effort because product reputation or advertising campaigns serve as a substitute for effort in the future, or the market approaches saturation. The results are robust to different forms of cost and demand externalities across customers. However, if the firm obtains intermediate information on sales rather than only aggregate sales figures, bonus plans dominate commissions even if the agent becomes harder to motivate over time. Information on the sales sequence allows the firm to pay bonuses specifically for success towards the end of the season, thereby overcoming the problem that the agent becomes demotivated if he cannot make quota anymore.

The analysis of optimal sales force compensation under moral hazard traces back to Basu et al. (1985). For a single-period setting with a risk-averse agent, they show that optimal incentive pay usually is a non-linear increasing function of sales. It is argued that commonly used compensation plans can be seen as a piecewise linear approximation of their optimal contract.² In contrast, I demonstrate that a simple linear incentive scheme can be the uniquely optimal compensation plan when dynamic aspects of sales force compensation are taken into account. This result obtains because the agent chooses effort several times before the firm can observe sales, and his incentive responsiveness may change over time. The analysis is closely related to Holmström and Milgrom (1987), who also analyze a multi-period principal-agent setting where the agent chooses effort after observing the history of previous outcomes. They show that

Demougine and Fluet (1998), Oyer (2000), Simester and Zhang (2010), Simester and Zhang (2014), Poblete and Spulber (2012), and Kishore et al. (2013).

²By now, there is an extensive literature analyzing optimal single-period sales force compensation in various contexts (see Coughlan (1993) and Albers and Mantrala (2008) for surveys). Dearden and Lilien (1990) and Lal and Srinivasan (1993) extend the work by Basu et al. (1985) to dynamic environments. Dearden and Lilien (1990) explain how commission rates should be adjusted in the presence of production learning effects. Lal and Srinivasan (1993) focus on a setting where, according to Holmström and Milgrom (1987), a linear incentive scheme is optimal.

linear contracts are optimal if the production technology is time- and history-independent and the agent exhibits stationary preferences as well as constant absolute risk aversion. In my model, the crucial departure from the Holmström-Milgrom framework is that the incentive problem can be time- and history dependent and the agent's preferences – in the form of his effort costs – can change over time. Otherwise, a linear contract is optimal in my model as well. In this sense, my findings are consistent with the Holmström-Milgrom framework and offer new insights on the optimality of linear incentive schemes when one departs from their model assumptions. In Holmström and Milgrom (1987), the principal does not benefit from obtaining more than an aggregate performance signal across periods. By contrast, in my setting, the optimality of a linear incentive scheme requires that the firm obtains only aggregate sales information.

The present paper is the first to characterize when combining a commission with a quota-based bonus is optimal. Advantages of bonus plans in dynamic settings have also been analyzed by, e.g., by Jain (2012) and Kishore et al. (2013).³ Focussing on a behavioral approach, Jain (2012) finds that multiperiod quotas can solve a self-control problem on the side of the agent. The current paper offers an alternative explanation for the optimality of bonus payments based on the rational behavior of sales agents. Kishore et al. (2013) show that quota-based bonuses exhibit an advantage over commissions when multitasking concerns are present, which is not an issue in my model.

The binary-effort approach employed in this paper is widely used in agency theory (e.g., Che and Yoo (2001); Laffont and Martimort (2002); Bolton and Dewatripont (2005); Schmitz (2005, 2013); Simester and Zhang (2010); Dai and Jerath (2013); Kaya and Vereshchagina (2014)). My framework is particularly related to dynamic binary-effort models with moral hazard and limited liability, which are studied by Bierbaum (2002), Bolton and Dewatripont (2005), and Schmitz (2005, 2013). Unlike the present paper, these authors all consider job design problems, assuming that the principal receives a performance signal every period. In this literature, it is common to assume that the principal always wishes to elicit high effort from the agent.⁴ This effort profile maximizes the principal's profit when his return in case of success is sufficiently large. I also take this approach, i.e., I focus on environments where the firm's revenue per sale is so large that the firm wants the agent to work hard with every customer. Kräkel and Schöttner (2014) endogenize the effort profile in a two-period model on optimal sales force compensation. In contrast to the present paper, in their model the agent is equally hard to incentivize in both periods. They show that, when the sale revenue is not large enough for a high-effort profile to

³Oyer (2000) shows that a bonus tied to a quota can be the uniquely optimal contract in a static setting.

⁴An exception is Schmitz (2005), who analyzes a two-period model where the principal implements low effort in the second period when his return in case of success is low.

be profit maximizing, a linear commission scheme is no longer optimal. The firm then either implements a pure quota-based bonus or a fixed salary.

The remainder of the paper is organized as follows. To explain the main dynamics of the model, I first solve a simple two-period problem, introduced in Section 2. Section 3.1 presents the solution to this model, which is extended to include cost externalities and demand externalities in Section 3.2. The impact of the information structure on the results is explored in Section 3.3, where I assume that the firm can observe sales each period. Section 4 analyzes the n -period case and Section 5 concludes.

2 The Basic Two-Period Model

A firm hires a sales agent to sell its product or service in each of two periods. In every period k ($k = 1, 2$), the sales agent can talk to one customer. To sell the product to the customer, the agent can exert low or high effort $e_k \in \{L, H\}$, $L < H$. Low effort may correspond to performing basic activities that can be easily monitored and thus enforced by the firm. For example, a firm can monitor whether a sales clerk is present at the shop to answer customer questions or to reload empty racks, or whether a sales representative contacts a customer. High effort may reflect that, in addition to the basic activities, the agent actively communicates the advantages of the firm's product in face-to-face encounters with customers or invests time and effort to learn a customer's specific needs. Such activities are usually prohibitively costly for the firm to monitor and, therefore, the firm cannot observe whether the agent provided the extra effort. I assume that the revenue per sale is sufficiently large so that the firm wants to induce high effort with every customer.

The probability that the period- k customer buys the product is α_k when effort is low and $\mu_k = \alpha_k + \rho_k$ when effort is high, with $\alpha_k, \rho_k > 0$ and $\mu_k < 1$. I exclude $\alpha_k = 0$ because in this case the firm can induce the efficient (first-best) solution without incurring agency costs. The parameter ρ_k reflects how responsive sales are to the agent's effort. The market for the product is large so that the agent's effort and sales in period 1 do not affect market characteristics in period 2. However, the probabilities α_k and ρ_k can vary exogenously. For example, the sales-effort responsiveness ρ_k could decrease over time because sales become less responsive to effort as customers get more familiar with the product. The agent incurs private costs for exerting effort. His costs for low effort are zero and his costs for high effort are $c_k > 0$ in period k , i.e., effort costs can also vary across periods. For example, the agent might get exhausted ($c_1 < c_2$) or there are learning effects ($c_1 > c_2$). In Section 3.2.1, I introduce the possibility that second-period cost depend on the first-period effort choice.

The firm neither observes the agent's effort choice nor his realized effort costs and thus encounters a typical moral-hazard problem. If the firm does not provide effort incentives, the agent will always exert low effort. Let $x_k \in \{0, 1\}$ indicate whether the agent sold the product ($x_k = 1$) or not ($x_k = 0$) in period k . At the end of period k , the agent observes x_k . The firm, however, only observes total sales $X_2 = x_1 + x_2$ at the end of period 2.⁵ It offers the agent a compensation plan that specifies a wage w_{X_2} for every possible outcome $X_2 = 0, 1, 2$. The firm designs the compensation plan to minimize expected wage costs for inducing high effort in each period.

To exclude trivial solutions to the given moral-hazard problem, I assume that the firm faces some contractual friction. Contract theory offers two standard frictions (e.g., Laffont and Martimort (2002), Sections 4.3 and 4.4) – the agent is assumed to be either risk averse (and unlimitedly liable) or protected by limited liability (and risk neutral). Both frictions imply that providing incentives leads to costs for the firm which exceed the effort and opportunity costs of the agent. In case of a risk-averse agent, the firm has to compensate him for any income risk he bears. In case of limited liability, the firm has to leave a rent to the agent in order to motivate him. In this paper, the sales agent is risk neutral⁶ but protected by limited liability in terms of $w_{X_2} \geq 0$ for all $X_2 = 0, 1, 2$. Accordingly, the firm cannot impose negative wages to punish the agent for poor performance. The agent's reservation value is zero. This assumption will imply that the firm has to leave a rent to the agent in the basic model.

The timeline is as follows. First, the firm offers a compensation plan w_{X_2} to the sales agent. The agent accepts or rejects the contract offer. If he rejects, the game will end and the agent earns his reservation value. If he accepts, he will choose effort $e_k \in \{L, H\}$ in every period k . At the end of the period, the agent observes x_k but the firm does not. At the end of period 2, the firm observes total sales X_2 and the agent is paid according to the compensation plan.

3 Optimal Compensation Plans with Two Periods

3.1 Solution to the Basic Model

Since the firm wants to induce high effort with every customer, we look for the cost-minimizing compensation plan that makes the agent prefer $e_k = H$ to $e_k = L$ in both periods $k = 1, 2$. I first derive the agent's incentive compatibility constraints, applying the common tie-breaking rule that the agent will choose high effort if he is indifferent between the two effort levels. In period

⁵The impact of this assumption on the results is discussed in Section 3.3.

⁶The empirical findings of Akerberg and Botticini (2002), Hilt (2008), and Bellemare and Shearer (2010) show that agents with low risk aversion sort themselves into risky jobs. Hence, it is not unrealistic to assume that sales agents, in particular, have a relatively high risk tolerance in practice.

2, the agent should exert high effort whether or not he sold the product to the first customer. The agent will do so if his expected payoff from choosing $e_2 = H$ exceeds his expected payoff from exerting effort $e_2 = L$ for each first-period outcome $x_1 \in \{0, 1\}$, i.e.,

$$\mu_2 w_{x_1+1} + (1 - \mu_2) w_{x_1} - c_2 \geq \alpha_2 w_{x_1+1} + (1 - \alpha_2) w_{x_1} \text{ for } x_1 = 0, 1.$$

These second-period incentive constraint can be rewritten as

$$w_2 - w_1 \geq R_2 \quad \text{and} \quad w_1 - w_0 \geq R_2, \tag{1}$$

where $R_k := \frac{c_k}{\rho_k}$ denotes the cost-responsiveness ratio in period k . Given that the agent chooses $e_2 = H$ in the second period, his expected payoff after a first-period success and failure are

$$W_s := \mu_2 w_2 + (1 - \mu_2) w_1 \quad \text{and} \quad W_f := \mu_2 w_1 + (1 - \mu_2) w_0,$$

respectively. Thus, the agent implements $e_1 = H$ in the first period if

$$\mu_1 W_s + (1 - \mu_1) W_f - c_1 - c_2 \geq \alpha_1 W_s + (1 - \alpha_1) W_f - c_2,$$

which is equivalent to

$$\mu_2 (w_2 - w_1) + (1 - \mu_2) (w_1 - w_0) \geq R_1. \tag{2}$$

According to the second-period incentive constraints (1), in order to motivate the agent in period 2, his compensation needs to increase by at least R_2 when he sells an additional unit of the product. In other words, $w_2 - w_1 = w_1 - w_0 = R_2$ are the smallest wage differences that induce high effort in period 2. From the first-period incentive constraint (2), it follows that a constant wage increase of R_2 per sale also induces high effort in period 1 if and only if $R_2 \geq R_1$. This condition states that the cost-responsiveness ratio is higher in the second than in the first period, implying that the agent's incentive responsiveness decreases over time. In this case, the smallest wage increase that is *necessary* to induce high effort in the second period is also *sufficient* to induce high effort in the first period. The linear compensation plan $w_0 = 0$, $w_1 = R_2$, and $w_2 = 2R_2$ then provides effort incentives at the lowest costs for the firm or, equivalently, the lowest rent to the agent.⁷ This compensation scheme, however, cannot induce high effort in the first period when $R_2 < R_1$. The firm should then also pay a bonus for the best possible sales outcome, as the first proposition shows.

⁷A complete proof that incorporates the agent's participation decision is given in the appendix. See the proof of Proposition 1.

Proposition 1 (i) If $R_2 \geq R_1$, the firm implements a pure commission scheme $w_{X_2} = R_2 \cdot X_2$.
(ii) If $R_2 < R_1$, the firm combines the commission with a bonus that is paid when the agent sold the product in both periods. The optimal compensation plan then is $w_0 = 0$, $w_1 = R_2$, and $w_2 = 2R_2 + \frac{1}{\mu_2}(R_1 - R_2)$.

All proofs are relegated to the Appendix.

Proposition 1 presents a key result of this paper that will be shown to extend to more general environments: The firm implements a linear incentive scheme – a pure commission – when the agent is hardest to incentivize in the last period. In such a situation, the firm should not pay a bonus for being successful in both periods because such a bonus plan is least effective when providing incentives is most crucial. The bonus fails to motivate the agent in the second period when he could not sell the product before. When R_2 is strictly larger than R_1 , the first-period incentive constraint (2) is not binding under the commission, implying that the firm overpays for providing first-period incentives. However, it is not possible to lower payments because then the agent will not always work hard in the second period.

If the agent is harder to incentivize in period 1 than in period 2, the commission R_2 still motivates the agent in the second period, but cannot induce high effort in the first period. If the firm wanted to maintain a pure commission scheme, it would have to raise the commission to R_1 in order to induce high effort in the first period. However, the firm would then always overpay for second-period incentives, independent of the first-period outcome. The firm can do better by complementing the commission R_2 with a bonus that is paid only when the agent was successful in both periods. Such a bonus is a powerful motivator in the first period when the agent always has the chance to make quota. The firm then overpays for second-period incentives only if the agent sold the product in the first period, thereby lowering the rent left to the agent. In other words, providing first-period incentives is more effectively accomplished by focussing rewards on the best-case sales scenario – as far as this is possible without demotivating the agent in case of a first-period failure.

According to Proposition 1, a linear compensation plan is optimal if $R_k = \frac{c_k}{\rho_k}$ is weakly increasing in k . Such a situation arises, for example, if the agent's effort costs are the same with every customer and the sales-effort responsiveness is time invariant because market characteristics do not change.⁸ The firm should also implement a pure commission when exerting high effort with the second customer is more costly than with the first (e.g., because dealing with customers exhausts the agent), while sales are less responsive to effort in the second period. Note that the latter does not necessarily mean that it becomes harder for the agent to sell the

⁸This result resembles the finding by Holmström and Milgrom (1987) that linear contracts are optimal if the production technology is not history-dependent and preferences are stationary.

product in the sense that a given customer is less likely to buy. Even though the sales-effort responsiveness ρ_k decreases in k , the sale probability under high effort, μ_k , can increase in k due to an increasing basic sale probability α_k . An increasing μ_k combined with a decreasing ρ_k can occur when the product is of particularly high quality and customers learn this over time, and therefore the agent's effort is less crucial for selling the product.

3.2 Robustness of the Results

3.2.1 Cost Externalities

I now demonstrate that the main insights from the basic model are robust to cost externalities across periods, which arise when the agent's effort choice in period 1 affects his effort costs in period 2. For example, interacting with a customer may be exhausting and more so the higher the agent's previous effort. Such negative cost externalities seem to be particularly relevant when customers arrive within a narrow time frame, e.g., when the agent works in a busy store as opposed to being a travelling salesman who can contact only one customer per day. Alternatively, learning effects might lead to positive cost externalities, i.e., working hard to sell the product to the first customer entails lower costs for dealing with a second one. To incorporate both types of externalities, I assume that the agent's costs for low effort remain zero in each period and the costs for high effort are still c_1 in period 1. The costs for high effort in period 2, however, are now given by

$$C_2 = c_2 + \Delta c_H \cdot I_{\{e_1=H\}} + \Delta c_L \cdot I_{\{e_1=L\}} > 0,$$

where $I_{\{\cdot\}}$ denotes an indicator variable that takes the value 1 if the statement $\{\cdot\}$ is true and the value 0 otherwise. Accordingly, if the agent exerts high (low) effort in the first period, his costs in the second period change by Δc_H (Δc_L) relative to his base costs c_2 . Negative cost externalities are reflected by $\Delta c_H \geq \Delta c_L \geq 0$. If in addition $c_2 \geq c_1$, the agent's effort costs are increasing and convex, reflecting a standard assumption in principal-agent models. However, we do not need to impose this as an assumption. Positive cost externalities correspond to $\Delta c_H \leq \Delta c_L \leq 0$ with $\Delta c_H < 0$. Dealing with a customer in the first period then lowers the costs of high effort in the second period, but total costs C_2 always remain positive.

Cost externalities add two effects to the basic model. First, they alter the agent's effort responsiveness. Given that the agent exerts high effort in the first period, the cost-responsiveness ratio in the second period changes to $R_2^c := \frac{1}{\rho_2}(c_2 + \Delta c_H)$. The new first-period cost-responsiveness ratio does not only include the cost increase from working hard in period

1, given by c_1 , but also the resulting cost externalities on the second period, $\Delta c_H - \Delta c_L$, and therefore becomes $R_1^c := \frac{1}{\rho_1}(c_1 + (\Delta c_H - \Delta c_L))$.⁹ Second, with sufficiently strong positive cost externalities, the firm does not need to leave a rent to the agent. Because the agent takes the positive cost externalities into account when he chooses effort in period 1, he is relatively easy to motivate in the first period. As a consequence, the lowest wages that induce high effort once the agent is locked into the contract may not cover the agent's expected costs ex ante. The firm then needs to pay a positive base wage to ensure the agent's participation. Neither of these two effects has an impact on the optimal structure of the compensation plan, which is characterized in Proposition 2.

Proposition 2 *Suppose that cost externalities exist. If the condition*

$$\mu_1 \max\{R_1^c, R_2^c\} + \mu_2 R_2^c \geq c_1 + (c_2 + \Delta c_H) \quad (3)$$

holds, the agent earns a rent and the optimal compensation plan is

$$w_{X_2}^c = \begin{cases} R_2^c \cdot X_2 & \text{if } R_2^c \geq R_1^c \\ R_2^c \cdot X_2 + \frac{1}{\mu_2} (R_1^c - R_2^c) \cdot I_{\{X_2=2\}} & \text{if } R_2^c < R_1^c \end{cases}.$$

If condition (3) does not hold, the firm minimizes its wage costs by complementing the compensation plan $w_{X_2}^c$ with a positive base wage.

The firm still employs a pure commission if the agent is harder to motivate in the second period and, otherwise, complements the commission with a bonus. If the agent does not earn a rent, the optimal compensation plan is no longer unique, as the proof of Proposition 2 shows.

3.2.2 Demand Externalities

So far I have assumed that the market is so large that first-period sales have no impact on future market characteristics. I now drop this assumption and introduce demand externalities across periods. In particular, I assume that it is not certain that there will be a customer in a given period because, e.g., no customer may pick up the phone, open the door, arrive at the shop, or is willing to talk to the sales agent. The firm cannot observe whether a customer arrives or not. The probability that a customer arrives in period k is denoted by $\beta_k \in (0, 1]$. The probability that a second customer arrives depends on the first-period sales outcome $x_1 \in \{0, 1\}$ so that $\beta_2 = \beta_2(x_1)$. Demand externalities arise if $\Delta\beta := \beta_2(1) - \beta_2(0) \neq 0$. A positive

⁹See the proof of Proposition 2 for a formal derivation of R_1^c and R_2^c based on the incentive constraints.

difference $\Delta\beta > 0$ indicates that first-period success increases the probability of a second-period customer compared to a first-period failure, e.g., because word-of-mouth advertising attracts new customers. A negative difference $\Delta\beta < 0$ can occur when the market has only few potential customers (“thin market”) so that a successful sale in the first period leads to a significant reduction of the remaining market capacity. This could be markets in which very expensive goods are traded like real estate markets or the high-end art market. If $\Delta\beta = 0$, there are no demand externalities but demand may still be uncertain in both periods. I further maintain the assumption of cost externalities.

Consider again the incentive compatibility constraints, starting with period 2. If there was a customer in the first period, the agent exerted high effort with this customer and thus his cost-responsiveness ratio in the second period is $R_2^c = \frac{1}{\rho_2}(c_2 + \Delta c_H)$. Otherwise, it remains $R_2 = \frac{c_2}{\rho_2}$ as in the basic model. The latter case can occur only if $\beta_1 < 1$. The second-period incentive constraints can thus be written as

$$w_2 - w_1 \geq R_2^c, \quad w_1 - w_0 \geq R_2^c, \quad \text{and} \quad w_1 - w_0 \geq R_2 \quad \text{if} \quad \beta_1 < 1. \quad (4)$$

Note that the last constraint is redundant for negative cost externalities but not for positive ones, which will crucially affect the optimal compensation plan. Now consider the first period and assume that a customer arrives. If the agent works hard, there will be a customer in the second period as well with probability $\beta^H = \mu_1\beta_2(1) + (1 - \mu_1)\beta_2(0) = \beta_2(0) + \mu_1\Delta\beta$. By contrast, if effort is low, a customer occurs in the second period with probability $\beta^L = \beta_2(0) + \alpha_1\Delta\beta$. The first-period incentive constraint becomes¹⁰

$$(\beta_2(0) + \Delta\beta) \cdot \mu_2(w_2 - w_1) + (1 - \beta_2(0)) \cdot \mu_2(w_1 - w_0) \geq R_1^d, \quad (5)$$

where

$$R_1^d := \frac{c_1 + (\beta^H \Delta c_H - \beta^L \Delta c_L) + \rho_1 \Delta\beta \cdot c_2}{\rho_1}$$

denotes the first-period cost-responsiveness ratio. R_1^d includes the expected cost externality on period 2, which now originates from two sources; the effect of high first-period effort on second-period effort costs and on the probability that a customer arrives in the second period. Proposition 3 derives the optimal compensation plan for negative cost externalities.

Proposition 3 *Suppose that there are negative cost externalities ($\Delta c_H \geq \Delta c_L \geq 0$) and de-*

¹⁰See the appendix for a derivation.

mand externalities. If the agent earns a rent,¹¹ the optimal compensation plan is

$$w_{X_2}^d = \begin{cases} R_2^c \cdot X_2 & \text{if } R_2^c \geq \frac{R_1^d}{1+\mu_2\Delta\beta} \\ R_2^c \cdot X_2 + \frac{1}{\beta_2(1-\mu_2)} [R_1^d - (1 + \mu_2\Delta\beta)R_2^c] \cdot I_{\{X_2=2\}} & \text{otherwise} \end{cases}.$$

If the agent does not earn a rent, the firm minimizes its wage costs by complementing the compensation plan $w_{X_2}^d$ with a positive base wage.

With negative cost externalities and demand externalities, the firm still implements a pure commission when the agent is sufficiently hard to incentivize in the second period. The corresponding threshold for the second-period cost-responsiveness ratio, however, does not longer equal the first-period cost-responsiveness ratio, now given by R_1^d . Instead, R_1^d has to be weighted by the degree of demand externalities $\Delta\beta$. This is because first-period effort now also affects the likelihood of a second customer and thereby the probability of being successful in both periods. This additional incentive effect is not included in R_1^d . If the commission R_2^c is too small to induce high first-period effort, the firm again additionally pays a bonus. Furthermore, due to demand externalities, the agent might not earn a rent.

Positive cost externalities in combination with short-term demand uncertainty ($\beta_1 < 1$) bring a new facet to the firm's contracting problem: The minimum wage differences $w_2 - w_1$ and $w_1 - w_0$ that ensure high effort in period 2 after a first-period success or failure, respectively, now differ. This is because a first-period failure can be due to the absence of a customer, in which case the agent's second period costs remain high. Formally, in the set of constraints (4), the second constraint is now redundant and the third one becomes relevant because $R_2 > R_2^c$. The wage difference $w_1 - w_0$ thus needs to be relatively large to ensure that the agent exerts high effort in period 2 even if he was not able to benefit from first-period learning effects. The wage difference $w_2 - w_1$, however, can be relatively small because two sales reveal to the firm that the agent was able to realize learning effects. The following proposition shows that, in such a situation, large second-period cost-responsiveness ratios no longer call for a pure commission. To keep the analysis tractable, I focus on a situation without demand externalities.¹²

Proposition 4 *Suppose that there are positive cost externalities ($\Delta c_H \leq \Delta c_L \leq 0$ and $\Delta c_H < 0$), short-term demand uncertainty ($\beta_1 < 1$), and no demand externalities ($\Delta\beta = 0$). The optimal compensation plan is given by $w_0 = 0$, $w_1 = R_2$, and $w_2 = 2R_2 + \max\left\{\frac{\Delta c_H}{\rho_2}, \frac{1}{\beta_2\mu_2} (R_1^d - R_2)\right\}$.*

¹¹The condition for this case to occur is stated in the proof of this proposition.

¹²With short-term demand uncertainty and positive cost externalities, the agent always earns a rent, as the proof of Proposition 4 shows. This might not be the case with demand externalities, which further complicates the analysis without adding any new insights.

The compensation plan characterized in Proposition 4 corresponds to a pure commission only in the special case where $R_2 = R_1^d$, i.e., when the agent is equally hard to incentivize in both periods. If $R_2 > R_1^d$, the firm pays a relatively small wage in case of two sales ($w_2 < 2R_2$) because it can then take advantage of the agent's increased incentive responsiveness due learning effects. The optimal marginal reward is thus decreasing. However, if $R_1^d > R_2$ and hence the agent is harder to motivate in period 1 than in period 2, the firm again combines a commission with a bonus to be paid for the best possible sales outcome.

Overall, the main insights from the basic model carry over to a situation with negative cost externalities and demand externalities. However, if positive cost externalities coincide with short-run demand uncertainty, a linear compensation scheme is in general not optimal. This result reveals that the broad optimality of pure commissions can break down if the agent's second-period incentive responsiveness depends on the arrival of a first-period customer.

3.3 Intermediate Information on Sales

In practice, firms often cannot monitor every encounter of the agent with a customer. Firms then obtain information on sales only after some time interval during which the agent might be able to talk to several customers. Accordingly, I have assumed that the firm only observes total sales. This assumption will now be dropped in order to discuss how it affects the results. I return to the basic two-period model as specified in Section 2, with the only difference that the firm now also learns the period- k sales outcome x_k ($k = 1, 2$). This allows the firm to condition wage payments on the sales sequence. Let $w_{x_1 x_2}$ denote the payment to the agent if he sold $x_k \in \{0, 1\}$ units of the product in period k . The second-period incentive constraints then become

$$\mu_2 w_{x_1 1} + (1 - \mu_2) w_{x_1 0} - c_2 \geq \alpha_2 w_{x_1 1} + (1 - \alpha_2) w_{x_1 0} \quad \text{for } x_1 = 0, 1$$

or equivalently

$$w_{11} - w_{10} \geq R_2 \quad \text{and} \quad w_{01} - w_{00} \geq R_2, \tag{6}$$

where $R_k = \frac{c_k}{\rho_k}$ still denotes the cost-responsiveness ratio in period k . Given that the agent chooses $e_2 = H$ in the second period, his expected payoff after a first-period success and failure now are

$$\hat{W}_s := \mu_2 w_{11} + (1 - \mu_2) w_{10} \quad \text{and} \quad \hat{W}_f := \mu_2 w_{01} + (1 - \mu_2) w_{00},$$

respectively. Thus, the agent works hard in period 1 if

$$\mu_1 \hat{W}_s + (1 - \mu_1) \hat{W}_f - c_1 - c_2 \geq \alpha_1 \hat{W}_s + (1 - \alpha_1) \hat{W}_f - c_2,$$

which is equivalent to

$$\mu_2(w_{11} - w_{01}) + (1 - \mu_2)(w_{10} - w_{00}) \geq R_1. \quad (7)$$

The second-period incentive constraints in (6) still require that the agent's wage increases at least by R_2 in response to a second-period sale. However, because it is possible to reward first- and second-period sales differently (i.e., $w_{10} \neq w_{01}$), the firm now has additional options to ensure that the first-period incentive constraint (7) holds. As a consequence, the optimal compensation plan is no longer unique, as the next proposition shows.

Proposition 5 *Assume that the firm obtains intermediate information on sales, i.e., it can observe x_1 and x_2 . An optimal compensation plan is then characterized by $w_{00} = 0$, $w_{01} = R_2$, and any wages w_{10} and w_{11} satisfying $w_{10} \in [0, R_1]$ and $w_{11} = R_2 + \frac{1}{\mu_2} R_1 - \frac{1 - \mu_2}{\mu_2} w_{10}$.*

The optimal wages in case of first-period failure, w_{00} and w_{01} , are uniquely determined. The firm, however, has infinitely many possibilities to compensate the agent for the outcomes that involve a first-period success. The following corollary discusses whether the compensation plans that are optimal if the firm can observe only total sales (compare Proposition 1) still belong to the class of optimal contracts.

Corollary 1 *Assume that the firm can observe x_1 and x_2 . A pure commission $w_{X_2} = R_2 \cdot X_2$ is optimal if and only if $R_1 = R_2$. Combining a commission with a bonus such that $w_{X_2} = R_2 \cdot X_2 + \frac{1}{\mu_2}(R_1 - R_2)I_{\{X_2=2\}}$ is optimal if and only if $R_1 > R_2$.*

When the firm can observe sales each period, a pure commission is optimal only in the special case where the agent is equally hard to incentivize in both periods. If the agent is harder to motivate in the first period, the firm still minimizes its wage costs by combining a commission with a bonus. Consequently, the firm benefits from intermediate information on sales only when the agent becomes more difficult to motivate over time ($R_2 > R_1$). In this case, the firm can lower the agent's rent when it conditions the compensation plan on the sales sequence instead of implementing a pure commission.

Corollary 2 *Assume that observing the sales outcomes x_1 and x_2 is costly. The firm benefits from collecting information on intermediate sales if and only if $R_2 > R_1$, i.e., if the agent becomes harder to incentivize over time.*

As discussed after Proposition 1, if only total sales are observable and $R_2 > R_1$, the firm overpays for first-period incentives. When the firm observes the sales sequence, it can overcome this problem by offering wages $w_{10} < w_{01}$, i.e., rewarding the agent less for a sale in period 1 than for a sale in period 2. Hence, the broad optimality of a linear compensation plan as described in Proposition 1 is driven by the assumption that the firm can only observe total sales. If the firm observes the sales sequence and $R_2 > R_1$, one optimal compensation plan is the combination of a commission with a bonus that is paid for a second-period sale.¹³ Hence, intermediate sales information can address the concern that a bonus scheme may not provide incentives in case of a first-period failure. The additional information allows the firm to pay the extra reward for a second-period sale independent of the first-period outcome.

4 Optimal Compensation Plans with n Periods

In this section, I extend the model with negative cost externalities from Section 3.2.1 to the general case of n periods. In each period k ($k = 1, \dots, n$), the sales agent can talk to one customer and chooses effort $e_k \in \{L, H\}$.¹⁴ The agent's costs for low effort are still zero in every period. The costs for high effort in period k are

$$C_k = c_k + \sum_{i=1}^{k-1} \Delta c_{iH} \cdot I_{\{e_i=H\}} + \sum_{i=1}^{k-1} \Delta c_{iL} \cdot I_{\{e_i=L\}}.$$

The first term, $c_k \geq 0$, denotes the agent's base costs in period k . The terms Δc_{iH} and Δc_{iL} with $\Delta c_{iH} \geq \Delta c_{iL} \geq 0$ characterize negative cost externalities across periods. Accordingly, if the agent exerts high (low) effort in period k , his costs for high effort increase by Δc_{kH} (Δc_{kL}) in all following periods. All other assumptions remain as in the basic model. In particular, only the agent observes whether he was successful in period k ($x_k = 1$) or not ($x_k = 0$) at the end of the period. The firm only observes total sales $X_n = \sum_{k=1}^n x_k$ and offers the agent a compensation plan w_{X_n} for $X_n = 0, 1, \dots, n$.

I again look for the cost-minimizing compensation plan that makes the agent prefer $e_k = H$ to $e_k = L$ in every period k and first derive the corresponding incentive compatibility constraints. Let $e = (e_1, \dots, e_n)$ denote the vector of the agent's effort choices. The firm wants to implement $e = e^H = (H, \dots, H)$. Consider an arbitrary period k and let X_{k-1} denote the total number of previously realized sales, i.e., $X_{k-1} := \sum_{i=1}^{k-1} x_i$. The agent chooses high effort instead of

¹³According to Proposition 5, one possible optimal compensation plan is $w_{10} = R_1$, $w_{01} = R_1 + B$, $w_{11} = 2R_1 + B$ with $B = R_2 - R_1$.

¹⁴The results derived in this section can be extended to include demand uncertainty in the sense that a customer arrives with an exogenously given probability $\beta_k \in (0, 1]$ in every period k .

low effort in period k if the associated expected wage increase meets or exceeds the expected increase in effort costs. The agent's final wage can be written as $w_{X_{k-1}+x_k+A_k}$ because total sales are the sum of the sales realized previous to period k , X_{k-1} , the period- k outcome x_k , and the sales realized after period k , denoted by $A_k := X_n - X_k$. Given that $e_j = H$ in all periods $j \neq k$, the agent chooses $e_k = H$ independent of his sales history if and only if

$$\begin{aligned} \rho_k \left(\sum_{a=0}^{n-k} \Pr[A_k = a | e = e^H, x_k = 1] \cdot w_{X_{k-1}+1+a} - \sum_{a=0}^{n-k} \Pr[A_k = a | e = e^H, x_k = 0] \cdot w_{X_{k-1}+a} \right) \\ \geq c_k + \sum_{i=1}^{k-1} \Delta c_{i_H} + \sum_{i=k+1}^n (\Delta c_{k_H} - \Delta c_{k_L}) \quad \text{for all } X_{k-1} = 0, \dots, k-1. \end{aligned} \quad (8)$$

The left-hand side of condition (8) corresponds to the increase in expected wage payments when the agent works hard. By exerting $e_k = H$, the agent increases his chance of selling the product in period k by ρ_k . If he sells the product in period k , his expected wage is given by the first sum in brackets. If he does not sell in period k , the second term in brackets gives his expected wage. The right-hand side of (8) characterizes the increase in expected effort costs: When choosing $e_k = H$, the agent incurs effort costs $c_k + \sum_{i=1}^{k-1} \Delta c_{i_H}$ (because he worked hard in all previous periods as well) and increases his future expected effort costs by $\sum_{i=k+1}^n (\Delta c_{k_H} - \Delta c_{k_L})$ relative to $e_k = L$. Constraint (8) needs to be satisfied for all periods $k = 1, \dots, n$.

In order to simplify (8), note that the probabilities α_i and ρ_i are independent of the sales history and we thus have

$$\begin{aligned} \Pr[A_k = a | e = e^H, x_k = 1] &= \Pr[A_k = a | e = e^H, x_k = 0] \\ &:= \Pr[A_k = a | e = e^H] \quad \text{for all } a = 0, \dots, n-k. \end{aligned} \quad (9)$$

Therefore, (8) holds for all $k = 1, \dots, n$ if and only if

$$\begin{aligned} \sum_{a=0}^{n-k} \Pr[A_k = a | e = e^H] (w_{X_{k-1}+1+a} - w_{X_{k-1}+a}) \geq R_k \\ \text{for all } X_{k-1} = 0, \dots, k-1 \text{ and } k = 1, \dots, n, \end{aligned} \quad (10)$$

where R_k again denotes the cost-responsiveness ratio in period k ,

$$R_k := \frac{1}{\rho_k} \left(c_k + \sum_{i=1}^{k-1} \Delta c_{i_H} + \sum_{i=k+1}^n (\Delta c_{k_H} - \Delta c_{k_L}) \right) \quad \text{for } k = 1, \dots, n.$$

The cost-responsiveness ratio characterizes the minimum expected wage increase due to a

period- k success that is necessary to induce high effort in period k . According to (10), in the last period $k = n$ it must hold that

$$w_{X_{n-1}+1} - w_{X_{n-1}} \geq R_n \text{ for all } X_{n-1} = 0, \dots, n-1.$$

Accordingly, to make the agent exert high effort in the last period for *any* number of previously realized sales X_{n-1} , his wage has to increase by at least R_n when he is successful in the last period. Therefore, the lowest wages that satisfy the last-period incentive constraints are $w_0 = 0$, $w_1 = R_n$, $w_2 = 2R_n$, and so on, which corresponds to the pure commission scheme $\bar{w}_{X_n} = X_n R_n$. If this commission also satisfies the constraints (10) for $k = 1, \dots, n-1$, it solves the firm's optimization problem. Note that $\sum_{a=0}^{n-k} \Pr[A_k = a | e = e^H] = 1$ in period k , because the probabilities in the sum correspond to all possible states of the world. Hence, substituting $w_{X_{k-1}+1+a} - w_{X_{k-1}+a} = R_n$ in (10) yields that $R_n \geq R_k$ is sufficient for \bar{w}_{X_n} being an optimal compensation plan.¹⁵ The following proposition also derives a condition under which $R_n \geq R_k$ is also necessary for a pure commission to be optimal.

Proposition 6 *The firm optimally implements a pure commission to induce high effort in all periods if the cost-responsiveness ratio is highest in the last period, i.e.,*

$$R_n \geq R_k \text{ for all } k = 1, \dots, n-1. \tag{C}$$

The optimal wages then are $\bar{w}_{X_n} = R_n X_n$. Moreover, if condition (C) does not hold and the sale probability in the last period is sufficiently high, i.e., $\mu_n \geq \frac{1}{2}$, then a pure commission is never optimal.

The first part of Proposition 6 extends result (i) from Proposition 1 and also has the same intuition. When condition (C) holds, the agent is hardest to incentivize in the last period. Therefore, a uniform wage increase of R_n per sale, that is necessary to induce high effort in the last period, makes the agent work hard in all other periods as well. The firm overpays for high effort in periods with cost-responsiveness ratios $R_k < R_n$ but cannot lower any payment because then the agent would not always work hard in the last period. When condition (C) does not hold and hence the agent is most difficult to motivate in some period $k^* < n$, the smallest pure commission that satisfies all incentive constraints in (10) and thus makes the agent work hard in all periods is $w_{X_n} = R_{k^*} X_n$. This commission, however, is not optimal if it is more likely than not that the agent will sell the product in the last period (i.e., $\mu_n \geq \frac{1}{2}$). The firm can

¹⁵The proof of Proposition 6 verifies that the agent's participation constraint holds for any incentive compatible compensation plan.

then lower the agent's rent by decreasing w_1 below R_{k^*} and increasing w_2 above $2R_{k^*}$. When the probability of selling the product in the last period is relatively high, the substitution rate w_2/w_1 , that maintains the agent's incentives in period k^* in case he has never been successful before, is relatively small. As a consequence, the firm's expected wage costs decrease when it shifts rewards from w_1 to w_2 .

When the agent's base effort cost c_k are increasing in k , then total effort costs C_k are increasing and convex across periods, as we might expect when customers arrive in a narrow time frame. Condition (C) holds and the firm implements a pure commission if, in addition, the sales-effort responsiveness is lowest in the last period. This is the case, e.g., when customers become more familiar with the product characteristics over time so that extra information provided by the sales agent is less crucial to sell the product.

Overall, Proposition 6 suggests that a commission is optimal in a broad class of settings. Three points are essential to obtain this result: First, the firm wants to induce high effort in every period irrespective of intermediate sales outcomes. If the firm was willing to forego high effort after the agent has failed to sell the product in one or more periods, a bonus plan might dominate a commission.¹⁶ Second, the agent is hardest to incentivize in the last period. Third, the market is large so that the sales history has no impact on the probability of sale, implying that $\Pr[A_k = a | e, x_k = 1] = \Pr[A_k = a | e, x_k = 0]$, which I used for the derivation of condition (C). Even if the third point is not satisfied, as it is for example the case under demand externalities, a very similar result may apply, as the analysis in Section 3.2.2 has shown.

According to part (ii) of Proposition 1, when there are two periods and a pure commission is not optimal, the firm complements the commission with a bonus that is paid for the most favorable sales outcome. The next proposition derives a condition under which such a compensation plan continues to be optimal in the n -period problem.

Proposition 7 *Suppose that $R_1 > R_n \geq R_k$ for all $k = 2, \dots, n - 1$. The optimal compensation plan then is*

$$w_{X_n} = R_n X_n + \frac{1}{\Pr[A_1 = n - 1 | e = e^H]} (R_1 - R_n) \cdot I_{\{X_n = n\}} \quad (\text{B})$$

That is, the firm pays the commission $\bar{w}_{X_n} = R_n X_n$ and an additional bonus if and only if the agent was successful in all periods.

Proposition 7 characterizes the optimal compensation plan for the special case where the agent is hardest to motivate in the first period and second hardest in the last period. The firm then combines the commission \bar{w}_{X_n} with a bonus for selling the product in all periods. The

¹⁶See Kräkel and Schöttner (2014) for a detailed discussion of this point for the case of two periods and a time-invariant incentive responsiveness of the agent.

optimal compensation plan can be derived by a backward induction argument, similar to the one that lead to Proposition 6. The wages w_1, \dots, w_{n-1} make the agent indifferent between working hard or not in the last-period. The wage w_n ensures that the first-period incentive constraint is just binding. The firm cannot benefit from increasing any wage w_j , $j = 1, \dots, n - 1$, above jR_n and simultaneously lowering w_n because concentrating rewards on the most favorable sales outcome provides first-period incentives at the lowest cost. The higher the number of periods, the lower is the probability that the agent will attain the bonus and, in particular, the agent cannot earn the bonus when he failed in the first period. However, this is not problematic from an incentive perspective because the prospect of a bonus payment only needs to motivate the agent in the first period. In the intermediate periods $2, \dots, n - 1$, the agent is easier to motivate than in the last period and, therefore, the commission induces high effort in all the remaining periods, even if the agent can no longer attain the bonus.¹⁷

One possibility to have $R_1 > R_n \geq R_k$ for all $k = 2, \dots, n - 1$ is that effort costs C_k are increasing across periods, but an increasing sales-effort responsiveness initially works against the effort costs effect. The sales-effort responsiveness is increasing if, with more experience, the agent gets better at his task in the sense that his extra effort becomes more effective. Alternatively, customer characteristics might change over time. For example, when the firm launches a new product, high effort might be less important to sell the product to early customers, because they are relatively more interested in the product and therefore better informed about its characteristics than customers who buy in later periods. We then obtain $R_1 > R_n \geq R_k$ for all $k = 2, \dots, n - 1$ if the sales-effort responsiveness initially increases substantially (due to a steep learning curve or changing customer characteristics), but finally the increase in effort costs dominates.

When the cost-responsiveness ratios R_k do not fall in the categories of Propositions 6 and 7, the firm's contracting problem becomes considerably more complex. In particular, a backwards induction argument as above can no longer be applied and optimal bonus plans are more variable in their structure. To illustrate this point, I now characterize the complete optimal compensation plan for the case $n = 3$. Using the incentive constraints (10) and the fact that $w_0 = 0$ is part of

¹⁷If the agent is risk averse and the number of periods large, such a bonus scheme requires the firm to pay a relatively high risk premium to the agent. A more balanced compensation plan may then be optimal.

the optimal compensation plan, the firm's problem for $n = 3$ is

$$\min_{w_1, w_2, w_3} \Pr[X_1 = 1|e = e^H] \cdot w_1 + \Pr[X_1 = 2|e = e^H] \cdot w_2 + \Pr[X_1 = 3|e = e^H] \cdot w_3 \quad (11)$$

$$\text{s.t. } \mu_2\mu_3(w_3 - w_2) + [\mu_2(1 - \mu_3) + \mu_3(1 - \mu_2)](w_2 - w_1) + (1 - \mu_2)(1 - \mu_3)(w_1 - w_0) \geq R_1, \quad (12)$$

$$\mu_3(w_2 - w_1) + (1 - \mu_3)(w_1 - w_0) \geq R_2, \quad (13)$$

$$\mu_3(w_3 - w_2) + (1 - \mu_3)(w_2 - w_1) \geq R_2, \quad (14)$$

$$w_1 - w_0 \geq R_3, \quad (15)$$

$$w_2 - w_1 \geq R_3, \quad (16)$$

$$w_3 - w_2 \geq R_3, \quad (17)$$

$$w_0 = 0. \quad (18)$$

The foregoing propositions already characterize the solution to this problem if the cost-responsiveness ratio is larger in the third than in the second period, $R_3 \geq R_2$. Proposition 6 applies if $R_3 \geq R_1$ and Proposition 7 otherwise. The following proposition therefore describes the optimal compensation plan for $R_2 > R_3$. Accordingly, it derives optimal wages for a decreasing cost-responsiveness ratio ($R_1 \geq R_2 > R_3$) and a situation where the agent is hardest to motivate in the second period ($R_2 \geq R_1$ and $R_2 > R_3$). The former case arises, e.g., when effort costs are time invariant and the sales-effort responsiveness increases across periods. The latter case might correspond to a situation where effort costs are increasing, and learning effects with respect to effort effectiveness take some time to kick in, but finally dominate the effort cost effect.

Proposition 8 *Suppose that $n = 3$ and $R_2 > R_3$. The firm never implements a pure commission. Defining $\Delta := \max\{R_1 - R_2, 0\}$, one of the following combined commission-bonus schemes is optimal.*

(i) *If $\frac{\Delta}{\mu_2} > \left(\frac{1}{\mu_3} - 2\right)(R_2 - R_3)$, the firm offers the wages*

$$\begin{aligned} w_0 &= 0, \quad w_1 = R_3, \quad w_2 = 2R_3 + \frac{1}{\mu_3}(R_2 - R_3), \\ w_3 &= 3R_3 + \frac{1}{\mu_3} \left(\frac{\Delta}{\mu_2} - \left(\frac{1}{\mu_3} - 3 \right) (R_2 - R_3) \right). \end{aligned} \quad (\text{P1})$$

The agent thus obtains a bonus if he sold two products, and an even higher bonus if he sold three products.

(ii) If $\frac{\Delta}{\mu_2} \leq \left(\frac{1}{\mu_3} - 2\right) (R_2 - R_3)$ and $\mu_1\mu_2\frac{(1-\mu_3)^2}{\mu_3^2} < 1$, the firm optimally offers

$$\begin{aligned} w_0 = 0, w_1 = R_3, w_2 = 2R_3 + \frac{1}{\mu_3} (R_2 - R_3), \\ w_3 = 3R_3 + \frac{1}{\mu_3} (R_2 - R_3). \end{aligned} \quad (\text{P2})$$

The firm thus pays a constant bonus if the agent sells at least two products.

(iii) If $\frac{\Delta}{\mu_2} \leq \left(\frac{1}{\mu_3} - 2\right) (R_2 - R_3)$ and $\mu_1\mu_2\frac{(1-\mu_3)^2}{\mu_3^2} \geq 1$, the firm pays the wages

$$w_0 = 0, w_1 = R_3 + D_1, w_2 = 2R_3 + D_2, \text{ and } w_3 = 3R_3 + D_2, \quad (\text{P3})$$

where D_1 and D_2 are functions of μ_2, μ_3, R_1, R_2 , and R_3 with $D_2 > D_1 > 0$. Accordingly, the agent obtains a bonus D_1 for one sale and a higher bonus D_2 for at least two sales.

Due to its length, the proof is relegated to an online appendix.

When there are three periods, the firm combines the commission R_3X_3 with a bonus scheme that can take three different forms. In case (i), the firm pays an additional bonus when the agent sold two or three products, and the bonus is higher in the latter case. In case (ii), a constant bonus for more than two sales is optimal. Finally, in case (iii), the firm pays a bonus already for one sale, and raises the bonus when there are at least two sales. Hence, compared to Proposition 7, we see that the firm may also pay a bonus when the agent was not successful in all periods. Furthermore, the compensation plans all comprise bonuses that are increasing in the number of sales, but not necessarily strictly increasing, as the cases (ii) and (iii) show. What bonus scheme is optimal depends on the relative size of the cost-responsiveness ratios, $R_1 - R_2$ and $R_2 - R_3$, and on the sale probabilities μ_1, μ_2 , and μ_3 . All these variables jointly determine which of the incentive constraints (12)-(17) are binding.

Case (i) occurs if and only if constraint (17) is *not* binding at the optimal solution, i.e., when an agent who has been successful in period 1 and 2 is relatively easy to incentivize in period 3. In case (i), the third-period incentive constraint (15) and the second-period constraint (13) are always binding, determining the optimal wages w_1 and w_2 , respectively. Given the optimal wages for one and two sales, the optimal w_3 follows either from the binding first-period constraint (12) or the second-period constraint (14). The former constraint is binding if and only if the agent is harder to incentivize in the first than in the second period, i.e., if $\Delta > 0$. Case (i) is more likely to occur if μ_3 is high; in particular, it always arises if $\mu_3 > \frac{1}{2}$. A high sale probability μ_3 makes it more worthwhile for the agent to be successful in period 2 and therefore w_2 can be low and incentive constraint (13) still holds.

In case (ii) and (iii), constraint (17) is always binding at the optimal solution, i.e., $w_3 = R_3 + w_2$. This implies that the agent's bonus does not increase when he sells three instead of two products. Unlike case (i), the issue is to ensure that $w_3 - w_2$ is large enough to make the agent exert high effort with the last customer when he has already sold the product twice. Starting from the (now infeasible) compensation plan (P1), the firm can achieve this in two different ways. First, it can keep w_1 and w_2 constant and just raise w_3 to $w_2 + R_3$, leading to a constant bonus payment for at least two sales (case (ii)). Second, the firm can simultaneously lower w_2 and increase w_1 , which entails an increasing bonus schedule (case (iii)). The constant bonus payment from case (ii) is optimal whenever the sale probabilities are identical in all periods $\mu_1 = \mu_2 = \mu_3 \leq \frac{1}{2}$. If the sale probabilities differ, case (iii) may occur if the first-period sale probability μ_1 is sufficiently high. To understand the intuition, observe that the agent's incentive constraints are all independent of μ_1 (holding ρ_1 constant). Hence, μ_1 only affects the firm's expected wage costs. The proof of Proposition 8 shows that the relative probability $\Pr[X_3 = 1]/\Pr[X_3 \geq 2]$ is decreasing in μ_1 . Hence, if μ_1 is high, raising w_1 above the lowest feasible payment R_3 can be worthwhile for the firm because it is relatively unlikely that the agent will be successful in only one period.

Overall, Proposition 8 shows that, if the agent is harder to motivate in period 2 than in period 3, the optimal compensation plan is not only determined by the relative size of the cost-responsiveness ratios R_k . The sale probabilities μ_k also play a crucial role. However, given that $\mu_1 = \mu_2 = \mu_3 \leq \frac{1}{2}$, a constant bonus payment for two and three sales is optimal whenever $R_1 < R_2$. The agent is then hardest to incentivize in the second period, which favors a strong wage increase when the agent sold the product at least twice.

5 Conclusion

This paper analyzes when a firm prefers a linear commission scheme to a quota-based bonus plan to motivate its sales force. In contrast to the sales agent, the firm only observes total sales after a certain time interval. The revenue per sale is sufficiently large so that the firm wants the agent to work hard with every customer. A pure commission is then optimal whenever the agent is most difficult to motivate in the last period. Otherwise, the firm can provide early-period incentives at lower cost by combining the commission with a bonus plan. The results are robust to different types of cost externalities and demand externalities. If, however, positive cost externalities and short-run demand uncertainty coincide, a linear compensation plan is optimal only if the agent's incentive responsiveness is time invariant. The reason is that, in this case, the wage difference required to motivate the agent in the last period crucially depends on

the previous arrival of customers. The results suggest that firms need to accurately investigate their specific environment before deciding on sales force incentives. The results derived in this paper can be tested empirically by using sales response functions to approximate the sales-effort responsiveness.

6 Appendix

Proof of Proposition 1. The firm's optimization problem reads as

$$\min_{w_0, w_1, w_2} \sum_{i=0}^2 \Pr[X_2 = i | e_1 = e_2 = H] \cdot w_i, \quad (19)$$

s.t. (1) and (2),

$$\sum_{i=0}^2 \Pr[X_2 = i | e_1 = e_2 = H] \cdot w_i - c_1 - c_2 \geq 0, \quad (20)$$

$$w_0, w_1, w_2 \geq 0. \quad (21)$$

The firm minimizes its expected wage costs subject to the agent's incentive compatibility constraints (1) and (2), participation constraint (20), and limited-liability constraints (21). I first show that (20) is implied by the remaining constraints and can therefore be neglected. Constraint (20) can be written as

$$W_f + \mu_1(W_s - W_f) \geq c_1 + c_2.$$

The constraints (1) and (21) together with $\mu_2 > \rho_2$ imply that $w_0 + \mu_2(w_1 - w_0) > c_2$, which is equivalent to $W_f > c_2$. Constraint (2) together with $\mu_1 > \rho_1$ implies that $\mu_1(W_s - W_f) > c_1$. Hence, the participation constraint (20) is non-binding and we can ignore it. Next, note that $w_0 = 0$ is optimal since a positive w_0 just lowers the agent's effort incentives while increasing the firm's wage costs. The firm's problem can thus be simplified to

$$\min_{w_1, w_2} \mu_1 \mu_2 \cdot w_2 + [\mu_1(1 - \mu_2) + \mu_2(1 - \mu_1)] \cdot w_1, \quad (22)$$

$$\text{s.t. } w_1 \geq R_2, \quad w_2 \geq R_2 + w_1, \quad w_2 \geq \frac{1}{\mu_2} R_1 - \frac{1 - 2\mu_2}{\mu_2} w_1. \quad (23)$$

This problem can be solved graphically in the w_1 - w_2 space. The first two constraints imply that the firm has to pay at least the wages $w_1 = R_2$ and $w_2 = 2R_2$. The firm therefore chooses these wages when they also satisfy the last constraint, which is the case iff $R_2 \geq R_1$. Hence, part (i) of Proposition 1 follows. Now consider the case $R_2 < R_1$. In the w_1 - w_2 -space,

the isocost line corresponding to (22) has a smaller slope than the last constraint because $-\frac{\mu_1(1-\mu_2)+\mu_2(1-\mu_1)}{\mu_1\mu_2} < -\frac{1-2\mu_2}{\mu_2}$ is equivalent to $\mu_2 > 0$ and thus holds. Hence, the optimal wages are $w_1 = R_2$ and $w_2 = \frac{1}{\mu_2}R_1 - \frac{1-2\mu_2}{\mu_2} \cdot R_2 = 2R_2 + \frac{1}{\mu_2}(R_1 - R_2)$, implying part (ii) of Proposition 1. ■

Proof of Propositions 2. Given that the agent works hard in the first period, the incentive compatibility constraints for period 2 are

$$\mu_2 w_{x_1+1} + (1 - \mu_2)w_{x_1} - (c_2 + \Delta c_H) \geq \alpha_2 w_{x_1+1} + (1 - \alpha_2)w_{x_1} \quad \text{for } x_1 = 0, 1.$$

Using $R_2^c = \frac{1}{\rho_2}(c_2 + \Delta c_H)$, they can be rewritten as

$$w_2 - w_1 \geq R_2^c, \quad w_1 - w_0 \geq R_2^c. \quad (24)$$

In the first period, the agent works hard iff

$$\mu_1 W_s + (1 - \mu_1)W_f - c_1 - (c_2 + \Delta c_H) \geq \alpha_1 W_s + (1 - \alpha_1)W_f - (c_2 + \Delta c_L).$$

This condition can be transformed to

$$\mu_2(w_2 - w_1) + (1 - \mu_2)(w_1 - w_0) \geq \frac{c_1 + (\Delta c_H - \Delta c_L)}{\rho_1} = R_1^c, \quad (25)$$

The firm's optimization problem thus is

$$\min_{w_0, w_1, w_2} \sum_{i=0}^2 \Pr[X_2 = i | e_1 = e_2 = H] \cdot w_i, \quad (26)$$

s.t. (24) and (25),

$$\sum_{i=0}^2 \Pr[X_2 = i | e_1 = e_2 = H] \cdot w_i - (c_1 + (c_2 + \Delta c_H)) \geq 0, \quad (27)$$

$$w_0, w_1, w_2 \geq 0. \quad (28)$$

The participation constraint (27) can be rewritten as

$$W_f + \mu_1(W_s - W_f) \geq c_1 + (c_2 + \Delta c_H). \quad (29)$$

I first consider negative cost externalities $\Delta c_H \geq \Delta c_L \geq 0$ and show that, in this case, (29) is implied by the remaining restrictions. To do so, I demonstrate that $W_f > c_2 + \Delta c_H$ and $\mu_1(W_s - W_f) > c_1$. The second constraint in (24) together with $\mu_2 > \rho_2$ and $w_0 \geq 0$ implies

that $w_0 + \mu_2(w_1 - w_0) > c_2 + \Delta c_H$, which is equivalent to $W_f > c_2 + \Delta c_H$. Constraint (25) is equivalent to $\rho_1(W_s - W_f) \geq c_1 + (\Delta c_H - \Delta c_L)$. Hence, because $\mu_1 > \rho_1$ and $\Delta c_H \geq \Delta c_L$, we obtain $\mu_1(W_s - W_f) > c_1$. The participation constraint is thus non-binding and can be neglected. It further follows that $w_0 = 0$ is optimal. The firm's problem can thus be simplified to

$$\min_{w_1, w_2} \mu_1 \mu_2 \cdot w_2 + [\mu_1(1 - \mu_2) + \mu_2(1 - \mu_1)] \cdot w_1, \quad (30)$$

$$\text{s.t. } w_1 \geq R_2^c, \quad w_2 \geq R_2^c + w_1, \quad w_2 \geq \frac{R_1^c}{\mu_2} - \frac{1 - 2\mu_2}{\mu_2} w_1. \quad (31)$$

This problem can be solved with the same solution procedure as problem (22)-(23) in the proof of Proposition 1. We obtain that a pure commission R_2^c is optimal iff $R_2^c \geq R_1^c$. If $R_1^c > R_2^c$, the firm pays the commission R_2^c and in addition a bonus $\frac{1}{\mu_2}(R_1^c - R_2^c)$ when the agent sold the product in both periods.

Now consider the case of positive cost externalities, $\Delta c_H \leq \Delta c_L \leq 0$ and $\Delta c_H < 0$. In contrast to the previous case, the participation constraint (29) is no longer implied by the remaining restrictions. Because $\Delta c_H \leq \Delta c_L$, we may no longer have $\mu_1(W_s - W_f) > c_1$. Therefore, the participation constraint can be binding. In order to solve the firm's problem, I proceed as follows. First, I solve a relaxed form of the firm's problem by dropping (27). Second, I derive the condition under which the solution to the relaxed problem also satisfies (27), implying that the agent still earns a rent under the optimal compensation plan. Afterwards, I consider the case where (27) is binding. The optimal solution to the relaxed problem again comprises $w_0 = 0$. Thus, the relaxed problem is identical to (30)-(31) and its solution is given by the compensation plan characterized above. The participation constraint (27) can be written as

$$\begin{aligned} w_0 + \Pr[X_2 \geq 1 | e_1 = e_2 = H] \cdot (w_1 - w_0) + \Pr[X_2 \geq 2 | e_1 = e_2 = H] \cdot (w_2 - w_1) \\ = w_0 + (\mu_1 + \mu_2 - \mu_1 \mu_2)(w_1 - w_0) + \mu_1 \mu_2 (w_2 - w_1) \geq c_1 + (c_2 + \Delta c_H). \end{aligned}$$

It is satisfied for the above compensation plans iff (3) holds.

Finally, assume that (3) does not hold, i.e., the participation constraint (27) is binding at the optimal contract. Thus, the agent's expected wage under an optimal compensation plan equals $\bar{C} = c_1 + (c_2 + \Delta c_H)$. The optimal compensation plan is no longer unique. All wages that satisfy (24) and (25) and lead to expected wage costs of \bar{C} are optimal. According to the previous analysis, one such compensation plan is as follows: If $R_2^c \geq R_1^c$, the firm implements the commission R_2^c and pays in addition a fixed wage $w_0 > 0$ that makes (27) binding, i.e., $w_0 = w_0^* := \bar{C} - (\mu_1 + \mu_2)R_2^c$, $w_1 = R_2^c + w_0^*$, $w_2 = 2R_2^c + w_0^*$. If $R_2^c < R_1^c$, the firm pays the

commission R_2^c and a bonus $\frac{1}{\mu_2}(R_1^c - R_2^c)$ if $X_2 = 2$ and a fixed wage $w_0 > 0$ that makes (27) binding, i.e., $w_0 = \hat{w}_0^* := \bar{C} - (\mu_1 R_1^c + \mu_2 R_2^c)$, $w_1 = R_2^c + \hat{w}_0^*$, $w_2 = 2R_2^c + \frac{1}{\mu_2}(R_1^c - R_2^c) + \hat{w}_0^*$. ■

Derivation of constraint (5). After a first-period success or failure, the agent's expected wage is

$$W_s^d := \beta_2(1)\mu_2 w_2 + (1 - \beta_2(1)\mu_2)w_1 \quad \text{and} \quad W_f^d := \beta_2(0)\mu_2 w_1 + (1 - \beta_2(0)\mu_2)w_0,$$

respectively. Hence, the agent's first-period incentive constraint is

$$\mu_1 W_s^d + (1 - \mu_1)W_f^d - (c_1 + \beta^H(c_2 + \Delta c_H)) \geq \alpha_1 W_s^d + (1 - \alpha_1)W_f^d - \beta^L(c_2 + \Delta c_L).$$

Using that $\beta^H - \beta^L = \rho_1 \Delta \beta$, the constraint can be rewritten as

$$W_s^d - W_f^d \geq \frac{c_1 + \rho_1 \Delta \beta \cdot c_2 + \beta^H \Delta c_H - \beta^L \Delta c_L}{\rho_1} = R_1^d$$

or

$$\beta_2(1) \cdot \mu_2(w_2 - w_1) + (1 - \beta_2(0)) \cdot \mu_2(w_1 - w_0) \geq R_1^d,$$

which is equivalent to (5). ■

Proof of Proposition 3. To shorten notation, define $\beta^{x_1} := \beta_2(x_1)$ for $x_1 \in \{0, 1\}$. I first derive the agent's expected effort costs C^d . Positive total effort costs arise in three different situations: (i) A customer arrived in both periods, leading to costs $c_1 + c_2 + \Delta c_H$, which happens with probability $\beta_1 \mu_1 \cdot \beta^1 + \beta_1(1 - \mu_1) \cdot \beta^0$. (ii) A customer arrived only in the first period, causing costs c_1 . This case arises with probability $(1 - \beta_1) \cdot \beta^0$. (iii) A customer arrived only in the second period, leading to costs c_2 , which happens with probability $\beta_1 \mu_1 \cdot (1 - \beta^1) + \beta_1(1 - \mu_1) \cdot (1 - \beta^0)$. Summing up and performing some transformations, using that $\beta^1 = \beta^0 + \Delta \beta$, gives

$$C^d = \beta_1 c_1 + (\beta^0 + \beta_1 \mu_1 \cdot \Delta \beta) c_2 + \beta_1 (\beta^0 + \mu_1 \cdot \Delta \beta) \Delta c_H.$$

The firm's optimization problem can then be written as

$$\min_{w_0, w_1, w_2} \sum_{i=0}^2 \Pr[X_2 = i | e_1 = e_2 = H] \cdot w_i, \quad (32)$$

s.t. (4) and (5),

$$\sum_{i=0}^2 \Pr[X_2 = i | e_1 = e_2 = H] \cdot w_i - C^d \geq 0, \quad (33)$$

$$w_0, w_1, w_2 \geq 0. \quad (34)$$

Due to demand externalities, the participation constraint (33) is not necessarily implied by the other constraints. I therefore first solve a relaxed version of the problem, neglecting constraint (33). I then check when the solution to the relaxed problem satisfies (33), corresponding to the case where the agent earns a rent. Neglecting (33), $w_0 = 0$ is optimal and the problem can be written as

$$\begin{aligned} \min_{w_1, w_2} & \Pr[X_2 = 1 | e = e^H] \cdot w_1 + \Pr[X_2 = 2 | e = e^H] \cdot w_2 \\ \text{s.t.} & \beta^1 \mu_2 \cdot w_2 + (1 - \mu_2 (\beta^0 + \beta^1)) \cdot w_1 \geq R_1^d, \\ & w_2 - w_1 \geq R_2^c, \\ & w_1 \geq R_2^c, \end{aligned}$$

This problem can be solved graphically in the $w_1 - w_2$ space. Defining $\gamma_1 := \beta_1 \mu_1$, the firm's objective function is equivalent to

$$\min_{w_1, w_2} W_f^d + \gamma_1 (W_s^d - W_f^d) = \min_{w_1, w_2} \gamma_1 \beta^1 \mu_2 w_2 + (\beta^0 \mu_2 + \gamma_1 (1 - \mu_2 (\beta^0 + \beta^1))) w_1.$$

The isocost line has a smaller slope than all the constraints iff

$$\begin{aligned} \frac{1 - \mu_2 (\beta^0 + \beta^1)}{\beta^1 \mu_2} & \leq \frac{\beta^0 \mu_2 + \gamma_1 (1 - \mu_2 (\beta^0 + \beta^1))}{\gamma_1 \beta^1 \mu_2} \\ \Leftrightarrow \gamma_1 (1 - \mu_2 (\beta^0 + \beta^1)) & \leq \beta^0 \mu_2 + \gamma_1 (1 - \mu_2 (\beta^0 + \beta^1)), \end{aligned}$$

which is true. Hence, the optimal wages are $w_1 = R_2^c$ and

$$\begin{aligned} w_2 &= \max \left\{ 2R_2^c, \frac{R_1^d}{\beta^1 \mu_2} - \frac{1 - \mu_2(\beta^0 + \beta^1)}{\beta^1 \mu_2} R_2^c \right\} \\ &= \max \left\{ 2R_2^c, \frac{R_1^d}{\beta^1 \mu_2} - \frac{1 - \mu_2(2\beta^1 - \Delta\beta)}{\beta^1 \mu_2} R_2^c \right\} \\ &= \max \left\{ 2R_2^c, 2R_2^c + \frac{R_1^d - (1 + \mu_2 \Delta\beta) R_2^c}{\beta^1 \mu_2} \right\} \end{aligned}$$

I now check when these wages satisfy the participation constraint (33). Constraint (33) can be written as

$$W_f^d + \gamma_1(W_s^d - W_f^d) \geq C^d.$$

We have $W_f^d = w_0 + \beta^0 \mu_2(w_1 - w_0)$ and $W_s^d - W_f^d = \beta^1 \mu_2(w_2 - w_1) + (1 - \beta^0 \mu_2)(w_1 - w_0)$. Hence, (33) can be further transformed to

$$w_0 + \beta^0 \mu_2(w_1 - w_0) + \gamma_1 [\beta^1 \mu_2(w_2 - w_1) + (1 - \beta^0 \mu_2)(w_1 - w_0)] \geq C^d.$$

Substituting for the optimal wages gives

$$\beta^0 \mu_2 R_2^c + \gamma_1 \left[\beta^1 \mu_2 \max \left\{ R_2^c, R_2^c + \frac{R_1^d - (1 + \mu_2 \Delta\beta) R_2^c}{\beta^1 \mu_2} \right\} + (1 - \beta^0 \mu_2) R_2^c \right] \geq C^d.$$

If this constraint holds, then the agent earns a rent and the optimal compensation plan is given by $w_0 = 0$, $w_1 = R_2^c$, and $w_2 = \left\{ 2R_2^c, 2R_2^c + \frac{R_1^d - (1 + \mu_2 \Delta\beta) R_2^c}{\beta^1 \mu_2} \right\}$. If the above condition is not satisfied, the agent's participation constraint is binding. The firm's optimal compensation plan is then no longer unique. Any plan that leads to expected wage costs of C^d and satisfies (24), (5), and (34) is optimal. For example, the firm can implement $w_1 = R_2^c$ and $w_2 = \left\{ 2R_2^c, 2R_2^c + \frac{R_1^d - (1 + \mu_2 \Delta\beta) R_2^c}{\beta^1 \mu_2} \right\}$ and pay a positive fixed wage w_0 that makes the participation constraint (33) just binding. ■

Proof of Proposition 4. The firm's optimization problem is

$$\min_{w_0, w_1, w_2} \sum_{i=0}^2 \Pr[X_2 = i | e_1 = e_2 = H] \cdot w_i, \quad (35)$$

s.t. (4) and (5),

$$\sum_{i=0}^2 \Pr[X_2 = i | e_1 = e_2 = H] \cdot w_i - (\beta_1 c_1 + \beta_2 (c_2 + \beta_1 \Delta c_H)) \geq 0, \quad (36)$$

$$w_0, w_1, w_2 \geq 0. \quad (37)$$

I first show that the participation constraint (36) is implied by the remaining restrictions. Constraint (36) can be rewritten as

$$W_f^d + \gamma_1(W_s^d - W_f^d) \geq \beta_1 c_1 + \beta_2(c_2 + \beta_1 \Delta c_H), \quad (38)$$

where $\gamma_1 = \beta_1 \mu_1$. Since $\beta_2 = \beta_2(0) = \beta_2(1)$, we have

$$W_s^d = \gamma_2 w_2 + (1 - \gamma_2) w_1 \quad \text{and} \quad W_f^d = \gamma_2 w_1 + (1 - \gamma_2) w_0,$$

with $\gamma_2 = \beta_2 \mu_2$. I now show that $W_f^d > \beta_2 c_2$ and $\gamma_1(W_s^d - W_f^d) > \beta_1(c_1 + \beta_2 \Delta c_H)$. The third constraint in (4), together with $\mu_2 > \rho_2$, implies that $\mu_2(w_1 - w_0) > c_2$. Multiplying both sides with β_2 and then adding $w_0 \geq 0$ to the left-hand side gives $w_0 + \gamma_2(w_1 - w_0) > \beta_2 c_2$, which is the same as $W_f^d > \beta_2 c_2$. From (5) and $\mu_1 > \rho_1$, we obtain $\mu_1(W_s^d - W_f^d) > c_1 + \beta_2(\Delta c_H - \Delta c_L)$, multiplying both sides with β_1 gives the desired inequality. Thus, $w_0 = 0$ and the firm's problem can be simplified to

$$\min_{w_1, w_2} \gamma_1 \gamma_2 \cdot w_2 + [\gamma_1(1 - \gamma_2) + \gamma_2(1 - \gamma_1)] \cdot w_1, \quad (39)$$

$$\text{s.t. } w_1 \geq R_2, \quad w_1 \geq R_2^c, \quad w_2 \geq R_2^c + w_1, \quad w_2 \geq \frac{R_1^d}{\gamma_2} - \frac{1 - 2\gamma_2}{\gamma_2} w_1. \quad (40)$$

Because $\Delta c_H < 0$, we have $R_2 > R_2^c$ and thus the constraint $w_1 \geq R_2^c$ can be dropped. Again using the same solution procedure as in the proof of Proposition 1, we obtain for the optimal compensation plan $w_1 = R_2$ and

$$w_2 = \max \left\{ R_2^c + R_2, \frac{R_1^d}{\gamma_2} - \frac{1 - 2\gamma_2}{\gamma_2} \cdot R_2 \right\} = 2R_2 + \max \left\{ \frac{\Delta c_H}{\rho_2}, \frac{1}{\gamma_2} (R_1^d - R_2) \right\}.$$

Thus, Proposition 4 follows. ■

Proof of Proposition 5. The firm's optimization problem is

$$\min_{w_{00}, w_{01}, w_{10}, w_{11}} \mu_1 \mu_2 w_{11} + (1 - \mu_1) \mu_2 w_{01} + \mu_1 (1 - \mu_2) w_{10} + (1 - \mu_1) (1 - \mu_2) w_{00}, \quad (41)$$

$$\text{s.t. (6) and (7),}$$

$$\mu_1 \mu_2 w_{11} + (1 - \mu_1) \mu_2 w_{01} + \mu_1 (1 - \mu_2) w_{10} + (1 - \mu_1) (1 - \mu_2) w_{00} - c_1 - c_2 \geq 0, \quad (42)$$

$$w_{00}, w_{01}, w_{10}, w_{11} \geq 0. \quad (43)$$

I first solve the problem neglecting the participation constraint (42). Afterwards I show that the solution of the relaxed problem satisfies (42). Neglecting (42), we must have $w_0 = 0$ and

$w_{01} = R_2$. Increasing the wages above these levels only increases the firm's wage costs while lowering the agent's first-period incentives, i.e., (7) is less likely to hold. The optimization problem thus simplifies to

$$\min_{w_{10}, w_{11}} \mu_1(\mu_2 w_{11} + (1 - \mu_2)w_{10}) + (1 - \mu_1)\mu_2 R_2,$$

$$w_{11} - w_{10} \geq R_2 \tag{44}$$

$$\mu_2 w_{11} + (1 - \mu_2)w_{10} \geq R_1 + \mu_2 R_2 \tag{45}$$

$$w_{10}, w_{11} \geq 0.$$

The firm thus wishes to minimize $\mu_2 w_{11} + (1 - \mu_2)w_{10}$. Hence, any wages $w_{10}, w_{11} \geq 0$ that satisfy (44) and make (45) binding are optimal. In other words, any non-negative wage pair satisfying $w_{11} \geq R_2 + w_{10}$ and $w_{11} = R_2 + \frac{1}{\mu_2}R_1 - \frac{1-\mu_2}{\mu_2}w_{10}$ is optimal. The largest w_{10} for which both constraints can be satisfied is $w_{10} = R_1$. It remains to show that these wages also satisfy (42). Plugging in $w_{00} = 0$ and $w_{01} = R_2$, constraint (42) can be rewritten as

$$\mu_1[\mu_2 w_{11} + (1 - \mu_2)w_{10}] + (1 - \mu_1)\mu_2 R_2 \geq c_1 + c_2.$$

Using that the term in square brackets is equal to $R_1 + \mu_2 R_2$, we obtain $\mu_1 R_1 + \mu_2 R_2 \geq c_1 + c_2$.

This is true because $R_k = \frac{c_k}{\rho_k}$ and $\mu_k > \rho_k$. ■

Proof of Corollary 1. By Proposition 5, an optimal compensation plan comprises $w_{01} = R_2$. Any commission-based compensation plan requires that $w_{10} = w_{01}$ and thus $w_{10} = R_2$. Because $w_{10} \in [0, R_1]$, $w_{10} = R_2$ is optimal iff $R_2 \leq R_1$. Assuming that $R_2 \leq R_1$ and setting $w_{10} = R_2$, we obtain $w_{11} = 2R_2 + \frac{1}{\mu_2}(R_1 - R_2)$. The corollary thus follows. ■

Proof of Proposition 6. The firm's optimization problem is

$$\min_{w_0, \dots, w_n} \sum_{i=0}^n \Pr[X_n = i | e = e^H] \cdot w_i \tag{46}$$

$$\text{s.t. (10)}$$

$$\sum_{i=0}^n \Pr[X_n = i | e = e^H] \cdot w_i - \sum_{i=1}^n \left(c_i + \sum_{j=1}^{i-1} \Delta c_{j_H} \right) \geq 0, \tag{47}$$

$$w_{X_n} \geq 0 \text{ for all } X_n = 0, \dots, n. \tag{48}$$

The firm's objective is to minimize the agent's expected wage payment subject to the set of incentive compatibility constraints (10), the participation constraint (47), and the limited liability constraints (48). I first show that (47) follows from the remaining constraints and can

thus be neglected. Let W_k^s (W_k^f) denote the agent's expected wage after success (failure) in period k , given that the agent did not sell the product in any previous period, i.e., $x_i = 0$ for all $i = 1, \dots, k-1$. The agent's expected wage can be written as

$$W_1^f + \mu_1(W_1^s - W_1^f) = W_2^f + \mu_2(W_2^s - W_2^f) + \mu_1(W_1^s - W_1^f) = \dots = W_n^f + \sum_{k=1}^n \mu_k(W_k^s - W_k^f).$$

Hence, using that $W_n^f = w_0$, constraint (47) is equivalent to

$$w_0 + \sum_{k=1}^n \mu_k(W_k^s - W_k^f) \geq \sum_{k=1}^n \left(c_k + \sum_{i=1}^{k-1} \Delta c_{iH} \right).$$

I next show that $\mu_k(W_k^s - W_k^f) > \left(c_k + \sum_{i=1}^{k-1} \Delta c_{iH} \right)$ for all $k = 1, \dots, n$. We have

$$\begin{aligned} W_k^s - W_k^f &= \sum_{a=0}^{n-k} \Pr[A_k = a | e = e^H, x_k = 1] w_{1+a} - \sum_{a=0}^{n-k} \Pr[A_k = a | e = e^H, x_k = 0] w_a \\ &\geq \frac{1}{\rho_k} \left(c_k + \sum_{i=1}^{k-1} \Delta c_{iH} \right), \end{aligned}$$

where the inequality follows from (8), using that $\Delta c_{kH} \geq \Delta c_{kL} \geq 0$. Because $\mu_k > \rho_k$ it follows that $\mu_k(W_k^s - W_k^f) > c_k + \sum_{i=1}^{k-1} \Delta c_{iH}$. Hence, because $w_0 \geq 0$, the participation constraint (47) is non-binding. Consequently, it is optimal to have $w_0 = 0$. From the analysis in the main text it follows that (C) is sufficient for a commission to be optimal and that the optimal commission then is \bar{w}_{X_n} . It remains to show that, if (C) does not hold and $\mu_n \geq \frac{1}{2}$, then a pure commission is not optimal. Assume that $k^* < n$ is the period where the agent is hardest to incentivize, i.e., $R_{k^*} = \max\{R_1, \dots, R_n\}$. To simplify the exposition, I assume that k^* is unique. However, in case it is not, the proof proceeds analogously. The lowest commission that satisfies all incentive constraints in (10) is R_{k^*} . I now assume that the commission R_{k^*} is an optimal compensation plan and lead this to a contradiction. If the commission R_{k^*} is optimal, all constraints in (10) with $R_k < R_{k^*}$ are non-binding. Thus, the commission R_{k^*} also solves the following relaxed and slightly transformed optimization problem:

$$\min_{W_1, \dots, W_n} \sum_{i=1}^n \Pr[X_n \geq i | e = e^H] \cdot W_i \quad (49)$$

$$\text{s.t. } \sum_{a=0}^{n-k^*} \Pr[A_{k^*} = a | e = e^H] W_{X_{k^*-1}+1+a} \geq R_{k^*} \text{ for all } X_{k^*-1} = 0, \dots, k^*-1, \quad (50)$$

where $W_j := w_j - w_{j-1}$ for $j = 1, \dots, n$ and $w_0 = 0$. Note that the wage difference W_1 appears in (50) only for $X_{k^*-1} = 0$, and that W_2 also appears in this constraint. Assume that W_1

is lowered by $\varepsilon > 0$. Constraint (50) continues to hold if W_2 is simultaneously increased by $\frac{\Pr[A_{k^*}=0|e=e^H]}{\Pr[A_{k^*}=1|e=e^H]}\varepsilon$. This change lowers the firm's expected wage costs if

$$\Pr[X_n \geq 2|e = e^H] \cdot \frac{\Pr[A_{k^*} = 0|e = e^H]}{\Pr[A_{k^*} = 1|e = e^H]}\varepsilon - \Pr[X_n \geq 1|e = e^H]\varepsilon < 0.$$

Because $\Pr[X_n \geq 2|e = e^H] < \Pr[X_n \geq 1|e = e^H]$, the last condition is satisfied if

$$\Pr[A_{k^*} = 1|e = e^H] \geq \Pr[A_{k^*} = 0|e = e^H]. \quad (51)$$

We have $\Pr[A_{k^*} = 1|e = e^H] \geq \mu_n \prod_{i=k^*+1}^{n-1} (1 - \mu_i)$. Hence, (51) holds if $\mu_n \prod_{i=k^*+1}^{n-1} (1 - \mu_i) \geq \prod_{i=k^*+1}^n (1 - \mu_i)$, which is true if $\mu_n \geq 1 - \mu_n$ or $\mu_n \geq \frac{1}{2}$. Hence, if $\mu_n \geq \frac{1}{2}$, a commission cannot be optimal. ■

Proof of Proposition 7. First, note that it is always optimal to set $w_0 = 0$ because, ceteris paribus, a positive w_0 only increases the firm's expected wage costs while making the incentive constraints (10) less likely to hold. Defining $W_j := w_j - w_{j-1}$ for $j = 1, \dots, n$, the firm's optimization problem can be rewritten as

$$\min_{W_1, \dots, W_n} \sum_{i=1}^n \Pr[X_n \geq i | e = e^H] \cdot W_i \quad (52)$$

$$\text{s.t. } \sum_{a=0}^{n-k} \Pr[A_k = a | e = e^H] W_{X_{k-1}+1+a} \geq R_k \text{ for all } X_{k-1} = 0, \dots, k-1 \text{ and } k = 1, \dots, n. \quad (53)$$

For simplicity, I write $\Pr[X_n \geq i] := \Pr[X_n \geq i | e = e^H]$ and $\sum_{a=0}^{n-k} \Pr[A_k = a] := \sum_{a=0}^{n-k} \Pr[A_k = a | e = e^H]$ in the following. I first solve a relaxed version of the above problem by neglecting all the constraint for $k = 2, \dots, n-1$ and the constraint for $k = n$ and $X_{k-1} = n-1$. The relaxed problem thus reads as

$$\min_{W_1, \dots, W_n} \sum_{i=1}^n \Pr[X_n \geq i] \cdot W_i \quad (54)$$

$$\text{s.t. } \sum_{a=0}^{n-1} \Pr[A_1 = a] W_{1+a} \geq R_1, \quad (55)$$

$$W_j \geq R_n \text{ for } j = 1, \dots, n-1. \quad (56)$$

I now show that $W_j = R_n$ for all $j = 1, \dots, n-1$ at the optimal solution of this problem. The proof is by contradiction. Assume that the optimal solution comprises a $\tilde{j} \in \{1, \dots, n-1\}$ such that $W_{\tilde{j}} > R_n$. I demonstrate that the firm can then lower its expected costs by decreasing $W_{\tilde{j}}$ and increasing W_n appropriately. Assume that the firm lowers $W_{\tilde{j}}$ by ε , where $0 < \varepsilon < W_{\tilde{j}} - R_n$.

Furthermore, to ensure that (55) still holds, the firm increases W_n by $\frac{\Pr[A_1=\tilde{j}-1]}{\Pr[A_1=n-1]}\varepsilon$. The expected wage costs decrease if and only if

$$\Pr[X_n \geq n] \frac{\Pr[A_1 = \tilde{j} - 1]}{\Pr[A_1 = n - 1]} \varepsilon - \Pr[X_n \geq \tilde{j}] \varepsilon < 0 \quad (57)$$

$$\Leftrightarrow \frac{\Pr[X_n \geq n]}{\Pr[A_1 = n - 1]} < \frac{\Pr[X_n \geq \tilde{j}]}{\Pr[A_1 = \tilde{j} - 1]} \quad (58)$$

$$\Leftrightarrow \frac{\mu_1 \mu_2 \cdots \mu_n}{\mu_2 \mu_3 \cdots \mu_n} < \frac{\Pr[X_n \geq \tilde{j} + 1] + \Pr[X_n = \tilde{j}]}{\Pr[A_1 = \tilde{j} - 1]}. \quad (59)$$

Using that $\Pr[X_n = \tilde{j}] = \Pr[A_1 = \tilde{j} - 1] \cdot \mu_1 + \Pr[A_1 = \tilde{j}] \cdot (1 - \mu_1)$, the last inequality becomes

$$\mu_1 < \frac{\Pr[X_n \geq \tilde{j} + 1] + \Pr[A_1 = \tilde{j}] \cdot (1 - \mu_1)}{\Pr[A_1 = \tilde{j} - 1]} + \mu_1,$$

which is true. Thus, we obtain that $W_j^* = R_n$ for $j = 1, \dots, n - 1$ at the optimal solution of the relaxed problem. The optimal W_n^* is then given by the binding constraint (55), i.e.,

$$\begin{aligned} & \Pr[A_1 = n - 1] W_n^* + (1 - \Pr[A_1 = n - 1]) R_n = R_1 \\ \Leftrightarrow W_n^* &= \frac{R_1 - (1 - \Pr[A_1 = n - 1]) R_n}{\Pr[A_1 = n - 1]} = R_n + \frac{1}{\Pr[A_1 = n - 1]} (R_1 - R_n). \end{aligned}$$

Because $W_j^* = R_n$ for $j = 1, \dots, n - 1$, $W_n^* > R_n$, and $R_n \geq R_k$ for all $k = 2, \dots, n - 1$, the solution W_1^*, \dots, W_n^* of the relaxed problem also satisfies the previously neglected constraints. Thus, W_1^*, \dots, W_n^* also solves the original problem. Substituting back, using $W_1 = w_1 = R_n$ and $W_j = w_j - w_{j-1}$ for $j = 2, \dots, n$, we obtain the optimal compensation plan (B). ■

References

- Akerberg, D. and M. Botticini (2002). Endogenous matching and the empirical determinants of contract form. *Journal of Political Economy* 110, 564–591.
- Albers, S. and M. Mantrala (2008). Models for sales management decisions. In B. Wierenga (Ed.), *Handbook of marketing decision models*. Springer: Berlin.
- Basu, A., V. S. R. Lal, and R. Staelin (1985). Salesforce compensation plans: An agency theoretic perspective. *Marketing Science* 4, 267–291.
- Bellemare, C. and B. Shearer (2010). Sorting, incentives and risk preferences: Evidence from a field experiment. *Economics Letters* 108, 345–348.
- Bierbaum, J. (2002). Repeated moral hazard under limited liability. Working Paper.

- Bolton, P. and M. Dewatripont (2005). *Contract Theory*. MIT Press: Cambridge, MA.
- Che, Y.-K. and S.-W. Yoo (2001). Optimal incentives for teams. *American Economic Review* 91, 525–541.
- Chung, D. J., T. Steenburgh, and K. Sudhir (2014). Do bonuses enhance sales productivity? a dynamic structural analysis of bonus-based compensation plans.
- Coughlan, A. T. (1993). Salesforce compensation: A review of MS/OR advances. In J. Eliashberg and G. L. Lilien (Eds.), *Handbooks in Operations Research and Management Science, Marketing*, Volume 5, pp. 611–651. Amsterdam: North-Holland.
- Dai, T. and K. Jerath (2013). Salesforce compensation with inventory considerations. *Management Science* 59, 2490–2501.
- Dearden, J. A. and G. L. Lilien (1990). On optimal salesforce compensation in the presence of production learning effects. *International Journal of Research in Marketing* 7, 179–188.
- Demougin, D. and C. Fluet (1998). Mechanism sufficient statistic in the risk-neutral agency problem. *Journal of Institutional and Theoretical Economics* 154, 622–639.
- Hilt, E. (2008). The negative trade-off between risk and incentives: Evidence from the american whaling industry. *Explorations in Economic History* 45, 424–444.
- Holmström, B. and P. Milgrom (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica* 55, 303–328.
- Innes, R. (1990). Limited liability and incentive contracting with ex-ante action choices. *Journal of Economic Theory* 52, 45–67.
- Jain, S. (2012). Self-control and incentives: An analysis of multiperiod quota plans. *Marketing Science* 31, 855–869.
- Joseph, K. and M. Kalwani (1998). The role of bonus pay in salesforce compensation plans. *Industrial Marketing Management* 27, 147–159.
- Kaya, A. and G. Vereshchagina (2014). Partnerships versus corporations: Moral hazard, sorting, and ownership structure. *American Economic Review* 104, 291–307.
- Kishore, S., R. Rao, O. Narasimhan, and G. John (2013). Bonuses versus commissions: A field study. *Journal of Marketing Research* 50, 317–333.

- Kräkel, M. and A. Schöttner (2014). Optimal salesforce compensation. *Working Paper Series of the Department of Economics, University of Konstanz 2014-09*.
- Laffont, J.-J. and D. Martimort (2002). *The Theory of Incentives: The Principal-Agent Model*. Princeton University Press: Princeton and Oxford.
- Lal, R. and V. Srinivasan (1993). Compensation plans for single- and multi-product salesforces: An application of the Holmstrom-Milgrom model. *Management Science* 39, 777–793.
- Oyer, P. (2000). A theory of sales quotas with limited liability and rent sharing. *Journal of Labor Economics* 18, 405–426.
- Poblete, J. and D. Spulber (2012). The form of incentive contracts: Agency with moral hazard, risk neutrality, and limited liability. *RAND Journal of Economics* 43, 215–234.
- Sappington, D. (1983). Limited liability contracts between principal and agent. *Journal of Economic Theory* 29, 1–21.
- Schmitz, P. (2005). Allocating control in agency problems with limited liability and sequential hidden actions. *RAND Journal of Economics* 36, 2, 318–336.
- Schmitz, P. W. (2013). Job design with conflicting tasks reconsidered. *European Economic Review* 57, 108–117.
- Simester, D. and J. Zhang (2010). Why are bad products so hard to kill? *Management Science* 56, 1161–1179.
- Simester, D. and J. Zhang (2014). Why do sales people spend so much time lobbying for low prices?
- Steenburgh, T. J. (2008). Effort or timing: The effect of lump-sum bonuses. *Quantitative Marketing and Economics* 6, 235–256.
- Zoltners, A. A., P. Sinha, and S. E. Lorimer (2008). Sales force effectiveness: A framework for researchers and practitioners. *Journal of Personal Selling and Sales Management* 28, 115–131.