Delegating Pricing Authority to Sales Agents: The Impact of Kickbacks

Matthias Kräkel* and Anja Schöttner†

Abstract

We analyze when a firm should delegate pricing authority to a sales agent who is better informed about the customer's valuation for the product than the firm. When the agent has pricing authority, customers may offer kickbacks to the agent to obtain a discount. The firm can prevent such collusion between agent and customer by designing the agent's performance pay appropriately, but may prefer not to do so. The reason is that potential kickbacks can motivate the agent to exert prospecting effort. We further study the optimal interaction between delegation and incentive pay.

Keywords: delegation, pricing authority, kickbacks, collusion, incentive pay

1 Introduction

Personal selling through sales forces is an important distribution channel for many firms (Zoltners et al., 2008). A long-standing question is whether and when firms should grant salespeople authority to set prices (Stephenson et al., 1979). A crucial advantage of delegating pricing authority is that salespeople are typically better informed about customers' willingness to pay than the firm and can thus optimally adapt the price to the customer's valuation for the product (Lal, 1986; Joseph, 2001).[1] On the downside, salespeople may

*University of Bonn, Institute for Applied Microeconomics, Adenauerallee 24-42, D-53113 Bonn, Germany, tel: +49 228 733914, e-mail: m.kraekel@uni-bonn.de.

†Humboldt-Universität zu 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]Frenzen et al. (2010) find empirical evidence that the degree of price delegation increases with the information asymmetry between the salesperson and the sales manager. Alternative explanations for price delegation are that it may have strategic commitment value that can soften price competition (Bhardwaj, 2001), or that it triggers positive reciprocity of salespeople (Lim and Ham, 2014).

abuse their pricing discretion at the expense of firm profits. Joseph (2001) theoretically shows that sales agents may rely on price discounts to boost sales rather than exert effort to identify high-valuation customers. Several empirical studies find that salespeople grant unnecessary price discounts to game the incentive system (e.g., Frank and Obloj, 2014; Larkin, 2014; Owan et al., 2015). Moreover, customers tend to bargain aggressively for price discounts when they know that a sales agent has pricing discretion (Stephenson et al., 1979). Sometimes customers and sales agents even agree to collude: Salespeople accept kickbacks from customers and in return grant a price discount even though the product could be sold at a higher price.[2]

This paper is concerned with a firm's problem of whether or not to delegate pricing authority to a sales agent who possesses superior knowledge about customers' valuations for the product, but can collude with customers. In order to identify a prospective customer, the agent has to exert non-observable search effort, which implies that the firm also faces a moral hazard problem.[3] Within an optimal contracting model, we characterize the circumstances under which the firm should delegate pricing authority and describe the optimal relationship between delegation and performance-based pay.

In our model, a prospective customer can have either a low or a high valuation for the firm's product, and the agent learns the valuation during the sales talk. When the customer's valuation is high but the agent is allowed to sell the product at a lower price, the customer will offer a kickback to the agent to get a price discount. The agent can accept the kickback and lower the price or reject the kickback and sell at a high price. Because it is typically easier for a sales agent to give a discount than to argue with the customer why there will not be one, we assume that the agent incurs private costs when he rejects a kickback (e.g., opportunity costs of time spent arguing). The firm can prevent collusion by means of contract design. It can either centralize pricing authority by stipulating a price that pertains to all customers, or, if it delegates pricing authority, reward the agent

[2]For example, in 2008 several employees of *Volkswagen Financial Services* were accused of selling used cars below market price to car dealerships and accepting bribes in return (see, e.g., "Schmiergeld-Vorwürfe gegen VW-Bank-Mitarbeiter", Süddeutsche Zeitung (2010, May 17), retrieved from http://www.sueddeutsche.de/wirtschaft/korruption-schmiergeld-vorwuerfe-gegen-vw-bank-mitarbeiter-1.202104). The guardian reports that an estate agent was caught offering properties to a developer below market price in return for bribes (see "Estate agent caught taking bribes", The Guardian (2004, May 4), retrieved from https://www. theguardian.com/business/2004/may/04/movinghouse.money).

[3]Prospecting is often seen as the most important activity of a sales agent (Weitz et al., 1998). In Section 5.1, we also consider a situation where the agent can exert effort to increase the customer's perceived valuation of the product during the sales talk.

for selling at a high price so that the agent is not susceptible to kickbacks. In addition to controlling the agent's pricing decisions, the contract also needs to motivate the agent to exert search effort, which requires rent payments to the agent because he is protected by limited liability.[4]

We show that the threat of collusion implies that the firm does not always want to delegate pricing authority. On the one hand, delegation allows the firm to utilize the agent's superior knowledge on how much the customer is willing to pay. On the other hand, it is costly to the firm to prevent the agent from abusing pricing authority: The agent's reward for selling at a high price needs to exceed the agent's benefit from colluding with the customer. Therefore, if it is sufficiently likely that a customer has a high willingness to pay, the firm centralizes pricing authority and stipulates a high price. It is then relatively unlikely that the firm will lose business because the agent faces a customer that is willing to pay only a low price.

However, the firm may also delegate pricing authority *precisely because* agent and customer can collude. Collusion decreases the price that the firm obtains for the product, but the firm benefits from lower incentive pay. The latter effect arises as the firm does not have to compensate the agent for arguing with the customer and because potential kickbacks already provide the agent with implicit incentives to search for a customer. These effects may dominate the negative price effect, and the firm then delegates pricing authority but does not prevent collusion by contract design. Allowing collusion is optimal when it is not very likely that a customer offers a kickback but if he does, the agent's arguing costs when rejecting the kickback are rather high. The latter implies that it would be too costly for the firm to prevent collusion. The firm's costs of collusion prevention are increasing in the agent's arguing costs and the potential kickback that a customer can offer, but decreasing in the effectiveness of the agent's search effort. When the probability of finding a prospective customer is highly responsive to search effort, the firm wants to elicit high effort from the agent and therefore implements high rewards for selling the product. Collusion can then be prevented as a byproduct of high-powered performance pay by rewarding the agent only if he sells the product at a high price.

A central question in organizational design is whether firms should accompany the delegation of decision rights with more or less incentive pay (Brickley et al., 2009). In

[4] Limited liability is a frequent assumption in the marketing literature (e.g., Simester and Zhang, 2010; Dai and Jerath, 2013) and very common in contract theory (e.g., Laffont and Martimort, 2002; Ohlendorf and Schmitz, 2012).

our model, the relationship between the delegation of pricing authority and the extent of performance pay is ambiguous and depends on whether or not the firm prevents collusion under delegation. When the firm allows for collusion, delegation is accompanied by lower incentive pay than centralization because delegation provides implicit incentives to the agent through kickbacks. By contrast, when the firm delegates but designs the contract such that the agent is not susceptible to collusion, delegation and performance pay can be positively related. The reason is that higher performance pay may be needed under delegation than under centralization to prevent the agent from abusing his authority, which is a standard argument for a complementary use of delegation and incentives also in other contexts (Holmström and Milgrom, 1991; Prendergast, 2002). Finally, our model predicts that the firm delegates pricing authority less often when it does not need to implement incentive pay to elicit effort from the agent. Such a situation may occur when sales agents can be closely monitored, e.g., because they work in-house and contact prospective customers by phone. When the agent does not need to be incentivized to exert effort, the rent that he earns under delegation due to the collusion problem is purely wasteful from the firm's perspective. As a consequence, delegation becomes less attractive in the absence of moral hazard.[5]

We contribute to the theoretical literature on the delegation of pricing authority under asymmetric information between the firm and the sales agent regarding the customers' willingness to pay (Lal, 1986; Joseph, 2001; Mishra and Prasad, 2004). To the best of our knowledge, our study is the first to consider the possibility of collusion between sales force and customers. Early contributions (Weinberg, 1975; Lal, 1986) emphasize that delegation is advantageous as long as the agent's incentives are properly aligned with the firm's objective. In contrast, Mishra and Prasad (2004) point out that centralized pricing is optimal if contracting occurs after the salesperson receives his private information, and Mishra and Prasad (2005) demonstrate that competitive product markets may favor centralized pricing. Our approach is closer to Joseph (2001), who also emphasizes the importance of search effort. In contrast to us, he models a trade-off between prospecting high-value and low-value customers and shows that the sales agent may substitute prospecting effort with charging low prices, which implies that full delegation of pricing may not be optimal. This also holds true in our model, where the firm always restricts the feasible price set to some extent. However, in contrast to Joseph (2001), optimal contracting may involve fully

[5] This result is in line with recent empirical evidence by Hong et al. (2015), who find that firms that adopt performance pay for exogenous reasons decentralize decision-making authority more often.

centralized pricing.

Theoretical and empirical studies on the optimal interaction between delegation of pricing authority and performance pay are still scarce.[6] The model by Joseph (2001) suggests a negative relationship between delegation and incentives, which we also predict for the case of delegation with collusion where the firm relies on implicit incentives from kickbacks. Lo et al. (2016) empirically study price delegation and performance pay for industrial equipment sales and find a positive relationship between delegation and incentives. This is in line with our results for the case where the firm optimally delegates pricing authority but prevents collusion. Lo et al. (2016) report that, in the context they study, sales people do not appear to "automatically" drop price when they are granted pricing authority, which could hint at the absence of collusion.[7]

As we address haggling for lower prices by high-valuation customers under price delegation, our paper is also related to the literature on price negotiation versus fixed-price policies (e.g., Riley and Zeckhauser, 1983; Wang, 1995; Desai and Purohit, 2004). According to this literature, the negotiation policy has the advantage that it allows for price discrimination but is also associated with different drawbacks, such as costs of haggling or hiring a larger sales force. We add to this literature by focusing on the role of sales agents in the selling process and studying the impact of price delegation on collusion and incentives to exert prospecting effort. We assume that haggling costs are incurred only by sales agents, and these costs are higher when a sales agent rejects a kickback than when he accepts one. This assumption reflects that a negotiation process is less uncomfortable for the agent and comes to an end more quickly when he immediately accepts the kickback offered by a customer.[8] Our model can be extended by assuming positive and heterogeneous haggling costs on the customer side (e.g., Desai and Purohit, 2004; Jindal and Newberry, 2015) which entails that only a fraction of high-valuation customers bargains for a lower price. The firm would then delegate and prevent collusion more often, because collusion prevention becomes less costly and implicit incentives from kickbacks decrease.

Finally, our paper is connected to the literature on collusion and supervision in organizations that was pioneered by Tirole (1986, 1992). In this literature, a firm wants to

[6]Papers that study optimal sales force compensation without the possibility of delegating pricing authority include Basu et al. (1985), Dearden and Lilien (1990), Lal and Srinivasan (1993), Kräkel and Schöttner (2016), and Schöttner (2016).

[7]Lo et al. (2016) also present a model. In contrast to our setting, the sales agent has to exert effort to learn the customer's valuation for the product and collusion is not an issue.

[8]For simplicity, we assume that haggling costs are zero when the sales agent accepts the kickback.

procure a product or service from an agent who has private information about his productivity or production costs. The firm can hire a supervisor who may be able to observe the agent's type and then makes a report to the firm. Our model could also be interpreted as a three-tier hierarchy where the party on the lowest hierarchy layer, the customer, is privately informed about the value of the relationship and the party at the second layer, the sales agent, can observe the agent's private information. However, our setting differs from the typical collusion model and we assume that the firm can neither communicate with the sales agent nor the customer, e.g., due to time constraints.[9] Hence, the firm cannot base its contract on reports by the sales agent or the customer. As common in the literature on price delegation, the only contracting variables are realized sales and price.

2 Basic Model

A firm wants to hire an agent to sell its product. We study a one-shot interaction between the agent and a single customer who buys at most one unit of the product. Before the agent can sell the product, he has to find a customer who has a positive valuation for the product, and conduct a sales talk. In order to find such a prospective customer, the agent needs to exert non-observable search effort $a \geq 0$. After having chosen a, the agent finds a customer with probability $g(a) \in [0, 1)$, satisfying $g(0) = 0$, $g'(a) > 0$, and $g''(a) \leq 0$. Exerting effort a leads to personal costs $c(a)$ for the agent (e.g., opportunity costs of time measured in monetary terms) with $c'(a), c''(a) > 0$ for $a > 0$ and $c(0) = c'(0) = 0$. We further assume that $\frac{c'(a)}{g'(a)}$ is convex in a.[10]

To allow for heterogeneous preferences, a customer's valuation (or willingness to pay) for the product, θ, can be either high, $\theta = \theta_H$, or low, $\theta = \theta_L$, with $\Delta\theta := \theta_H - \theta_L > 0$ and $\theta_L > 0$. Production costs are smaller than θ_L so that the firm always wants to sell the product to a prospective customer. For simplicity, we normalize production costs to zero. Before a customer is found, the firm and the agent have the common knowledge that with

[9]Simester and Zhang (2014) study internal lobbying by sales people and analyze a setting where the sales agent reports back to the firm after observing the customer's type. The application they have in mind is business-to-business settings where firms engage in extended sales processes with large customers. Our setting applies to selling standardized products that will be sold (or not) to a customer without getting back to the firm (e.g., selling perishable products, retail).

[10]A sufficient condition for the convexity of $\frac{c'(a)}{g'(a)}$ is that $c'''(a) \geq 0$ and $g'''(a) \leq 0$. Convexity of $\frac{c'(a)}{g'(a)}$ is also given for $g(a) = a^\rho$ and $c(a) = \frac{a^2}{2c}$, with parameters $\rho \in (0, 1]$ and $c > 0$, which we will introduce in Section 4 as specific functional forms.

probability $q \in (0, 1)$ a customer has a high valuation, θ_H, and with probability $1 - q$ a customer has a low valuation, θ_L. For example, the firm and the agent may know from market research data that the share of θ_H-customers in the market amounts to q, and the share of θ_L-customers is given by $1 - q$. When the agent has found a customer, he learns the customer's valuation θ during the sales talk. By contrast, the firm never observes θ. This assumption reflects the typical informational advantage of sales agents who directly contact customers.

With respect to the customer, we make the rather general assumption that he knows θ before he makes his first decision in the sales process, which implies that he may learn his valuation at different points in time. For example, the customer may know his valuation θ before being contacted by the agent, or the customer may learn about important product properties only during the sales talk so that his personal valuation θ is realized after being approached by the agent. As a third alternative, θ could also describe the customer's posterior expected valuation of the product after the sales talk, implying that his exact preferences remain uncertain at the time of sale.[11]

For contracting purposes, it is only verifiable whether the agent has sold the product or not and, in case of a sale, what price the customer has paid.[12] Thus, the firm specifies incentive pay $(w_\varnothing, w(p))$ for the agent, with $w_\varnothing$ denoting the wage paid to the agent if no sale is realized and $w(p)$ denoting the wage if the product is sold at price p. Furthermore, the firm stipulates a feasible price set P, which means that the agent is allowed to offer only prices $p \in P$ to the customer. If P contains only a single price, the agent has no pricing authority (*centralization of pricing authority*). Otherwise, the agent has discretion over the price (*delegation of pricing authority*). The wage schedule $w(p)$ is stipulated for all $p \in P$. To sum up, the agent's contract is given by $(w_\varnothing, w(p), P)$. Firm and agent are assumed to be risk neutral.[13] Furthermore, the agent is protected by limited liability so that wages have to be non-negative, and his reservation value is zero.

We design our model to capture two key aspects of delegation. On the one hand,

[11] This is in line with typical assumptions of the advertising literature (e.g., Nelson, 1970, 1974; Anderson and Renault, 2006; Bagwell, 2007). For example, if the good is a search good, customers either know their valuations in advance or learn them during the sales talk. In case of an experience good, customers do not know their exact valuations but form different beliefs and posterior expected valuations during the talk.

[12] Hence, if the product was not sold, it is not verifiable whether the agent has found a prospective customer to conduct a sales talk or not.

[13] The empirical studies by Ackerberg and Botticini (2002), Hilt (2008), and Bellemare and Shearer (2010) document that, typically, less risk averse and risk neutral individuals sort themselves into more risky jobs, e.g., as a sales agent, whose income often depends on realized sales and is, hence, quite risky.

delegation allows utilizing the agent's knowledge on the customer's valuation in the price setting process. On the other hand, the customer could try to obtain a price discount when he is aware that the agent has pricing authority. In practice, customers may learn from word of mouth, own observations, or the firm's advertised pricing policy whether sales agents of a specific firm typically grant price discounts or not. In our model, we assume for simplicity that the feasible price set P is publicly observable. When the agent makes a price offer $\tilde{p}$ to a customer who is aware that a lower price is feasible, the customer offers a kickback (e.g., bribe, tip, or favor) to the agent for granting a price discount. The agent then has two options. He can either accept the kickback and lower the price, or he can reject the kickback and insist on his initial price offer $\tilde{p}$. In the latter situation, the agent incurs private arguing costs $\kappa \geq 0$ from a time-consuming discussion with the customer.[14]

The precise rules and the timing of the game are as follows.

1. The firm specifies a contract $(w_\varnothing, w(p), P)$ for the agent, where p_L (p_H) denotes the lowest (highest) price in the feasible price set P.

2. The agent accepts or rejects the contract $(w_\varnothing, w(p), P)$. In the latter case, the game ends. If the agent accepts, the game proceeds.

3. The agent chooses search effort a at cost $c(a)$ and finds a customer with probability $g(a)$. If, with probability $1 - g(a)$, no customer is found, the agent will receive $w_\varnothing$ and the game ends. If a customer is found, the game proceeds.

4. The agent conducts a sales talk and learns the customer's valuation of the product, $\theta \in \{\theta_L, \theta_H\}$.

5. The agent chooses a price offer $\tilde{p}$ such that $\tilde{p} \in P$ and $\tilde{p} \leq \theta$. If no such $\tilde{p}$ exists, the agent will receive $w_\varnothing$ and the game ends. Otherwise, if $\tilde{p} = p_L$, the customer buys the product at price p_L, the agent obtains $w(p_L)$ and the game ends. If $\tilde{p} > p_L$, the customer offers the kickback $\beta \cdot (\tilde{p} - p_L)$, with $\beta \in (0, 1)$, in order to buy the product at the price of p_L.

6. The agent decides whether to accept or reject the kickback. If he accepts, his payoff will be $w(p_L) + \beta(\tilde{p} - p_L)$. If he rejects, his payoff will be $w(\tilde{p}) - \kappa$.

[14] See the introduction for more detailed explanations.

At stage 5, in case the firm has decided to centralize pricing authority, we have $P = \{p_L\}$. The agent will hence offer the price $\tilde{p} = p_L$ if this price does not exceed the customer's valuation. The customer, being aware that the agent cannot grant a price discount, does not offer a kickback. By contrast, under delegation, the agent knows that the maximal feasible price discount is $\tilde{p} - p_L$. The offered kickback $\beta(\tilde{p} - p_L)$ depends on a parameter β that may reflect relative bargaining power or inefficiencies in the bargaining process (e.g., the customer may offer a favor to the sales agent that is more costly to the customer than it is valued by the agent). The agent and the firm cannot communicate before the sale is closed, e.g., because of lack of time.

As tie-breaking rules, we assume that if the agent is indifferent between offering different prices, he is loyal to the firm and offers the highest of these prices. If the agent is indifferent between accepting and rejecting a kickback, he will reject it. If the customer is indifferent between buying and not buying the product, he will buy.

3 Possible Contract Types

A fundamental contractual choice that the firm makes in our model is whether to centralize or delegate pricing authority. In this section, we derive the *conditionally optimal* contracts *given that* the firm implements either centralization (i.e., P is a singleton) or delegation (i.e., P contains at least two elements). Building on these results, we analyze when the firm prefers delegation to centralization and vice versa in Section 4.

3.1 Centralization of Pricing Authority

When the firm centralizes pricing authority, it optimally chooses either $P = \{\theta_L\}$ or $P = \{\theta_H\}$. Any price $p > \theta_H$ would prevent a sale with certainty, whereas any price $p < \theta_L$ would leave an unnecessary rent to the customer. In addition, any price $p \in (\theta_L, \theta_H)$ is dominated by the price $p = \theta_H$, which also leads to a sale only with a θ_H-customer but without leaving a rent to him. We henceforth denote a centralization contract that stipulates $P = \{\theta_L\}$ as a contract of type C_L. By contrast, a centralization contract specifying $P = \{\theta_H\}$ is a contract of type C_H. In the next step, we determine the firm's optimal incentive pay for each contract type.

First, suppose the firm chooses contract type C_L, i.e., $P = \{\theta_L\}$. The agent then sells to both customer types at price θ_L and earns the wage $w_L := w(\theta_L)$. Hence, when designing the optimal incentive pay that complements $P = \{\theta_L\}$, the firm solves

$$\max_{w_\varnothing, w_L} g(a) \cdot (\theta_L - w_L) - (1 - g(a)) \cdot w_\varnothing \quad \text{subject to}$$

$$a \in \arg\max_{\hat{a}} g(\hat{a}) \cdot w_L + (1 - g(\hat{a})) \cdot w_\varnothing - c(\hat{a}), \tag{1}$$

$$g(a) \cdot w_L + (1 - g(a)) \cdot w_\varnothing - c(a) \geq 0, \tag{2}$$

$$w_\varnothing, w_L \geq 0. \tag{3}$$

The firm maximizes expected net profits subject to three constraints. Constraint (1) describes the agent's incentive constraint, i.e., for given wages $w_\varnothing$ and w_L the agent chooses the level of search effort, a, that maximizes his expected net income. The participation constraint (2) requires that the agent's expected net income must be at least as large as his zero reservation value. The limited-liability constraint (3) restricts wages to non-negative values. As $c(0) = 0$, the limited-liability constraint (3) implies that the agent can always ensure himself a non-negative expected net income by choosing zero effort. Hence, the limited-liability constraint (3) already implies the participation constraint (2), which can therefore be neglected in the following. Furthermore, the agent's wage for not selling the product, $w_\varnothing$, increases the firm's labor costs and decreases the agent's search incentives (see (1)). Hence, the limited-liability constraint is binding for $w_\varnothing$ so that $w_\varnothing = 0$. Finally, as the agent's objective function is strictly concave, the incentive constraint (1) can be replaced by the first-order condition $w_L = c'(a)/g'(a)$. Altogether, the firm solves

$$\max_{w_L \geq 0} g(a)(\theta_L - w_L) \quad \text{subject to} \quad a = e(w_L),$$

with $e(\cdot)$ as the (monotonically increasing and concave) inverse function of $c'(a)/g'(a)$.

Next consider the case where the firm chooses contract type C_H, i.e., $P = \{\theta_H\}$. Now the agent can sell the product only to θ_H-customers because the price exceeds the valuation of θ_L-customers. Let $w_H := w(\theta_H)$ denote the agent's wage when a sale is realized, which happens with probability q. In strict analogy to contract type C_L, to determine the optimal incentive pay the firm solves

$$\max_{w_H \geq 0} g(a) \cdot q \cdot (\theta_H - w_H) \quad \text{subject to} \quad a = e(qw_H).$$

In order to compare the two contract types C_L and C_H, it is useful to rewrite the optimization problems such that the firm chooses the agent's expected compensation W. To this end, define $A(W) := g(e(W))$ as the probability that the agent finds a customer when the expected wage is W.[15] The firm chooses W to solve

$$\max_{W \geq 0} A(W)(\Theta - W) \quad \text{with} \quad (\Theta, W) := \begin{cases} (\theta_L, w_L) & \text{if} \quad P = \{\theta_L\} \\ (q\theta_H, qw_H) & \text{if} \quad P = \{\theta_H\}. \end{cases} \tag{4}$$

Let $W^*(\Theta)$ denote the optimal solution to this problem, which is implicitly described by

$$\Theta = \frac{A(W^*(\Theta))}{A'(W^*(\Theta))} + W^*(\Theta). \tag{5}$$

Objective function (4) immediately shows that the firm will prefer $P = \{\theta_L\}$ to $P = \{\theta_H\}$ if and only if $\theta_L \geq q\theta_H$ and then pays the wage $w_L^* := W^*(\theta_L)$. In case of $\theta_L < q\theta_H$, the firm stipulates $P = \{\theta_H\}$ and pays the wage $w_H^* := W^*(q\theta_H)/q$. Accordingly, if it is sufficiently unlikely that a customer has a high valuation for the product, the firm will prefer contract type C_L to contract type C_H. Otherwise, the firm chooses contract type C_H although the high product price prevents a sale when the customer has a low valuation for the product. The following proposition summarizes our findings:

Proposition 1 *Suppose the firm implements centralization. It will choose contract type C_L and pay the wage w_L^* for a sale if and only if $\theta_L \geq q\theta_H$; otherwise, the firm chooses contract type C_H and pays the corresponding wage w_H^* for a sale. The optimal wages w_L^* and w_H^* are implicitly described by*

$$\theta_L = \frac{A(w_L^*)}{A'(w_L^*)} + w_L^* \quad \textit{and} \quad q\theta_H = \frac{A(qw_H^*)}{A'(qw_H^*)} + qw_H^*. \tag{6}$$

3.2 Delegation of Pricing Authority

We now turn to the analysis of delegation, where the feasible price set P has at least two elements. The following lemma facilitates the further analysis.[16]

Lemma 1 *Under delegation, the analysis can without loss of generality be restricted to the case where the firm chooses $P = \{\theta_L, \theta_H\}$ as feasible price set.*

[15] Note that $A(W)$ is concave.

[16] See the Supplementary Material for a proof.

According to Lemma 1, the firm optimally chooses the two possible customer valuations as the feasible prices that the agent can offer. To understand the intuition, recall that the firm has two options under delegation. First, the firm can prevent collusion through an appropriate contract design. In this case, stipulating a maximum feasible price p_H that is below θ_H has a potential advantage. It would force the agent to propose lower prices, which in turn lead to lower kickback offers by the customer. Hence, collusion would become less attractive to the agent and could therefore be easier prevented by the firm. However, this effect is always dominated by the firm's loss due to obtaining a lower price from a θ_H-customer. Second, the firm can allow for collusion and use kickbacks as implicit effort incentives for the agent. Then, specifying a minimum feasible price p_L that is below θ_L would increase kickbacks and hence implicit incentives. However, the incentive effect is always second order in comparison to the firm's loss from having customers pay a lower price for the product.[17] These results imply that the firm should never grant full pricing authority where no restrictions on P are imposed, as already suggested by Stephenson et al. (1979) and also found in the theoretical model by Lo et al. (2016).

Consider the agent's behavior under a given contract $(w_\varnothing, w(p), \{\theta_L, \theta_H\})$ when he has found a customer. If the agent learns that the customer has a low valuation of the product, then the agent offers the price $\tilde{p} = \theta_L$ and the customer buys at this price. When the agent learns that the customer has a high valuation of the product, he can either offer $\tilde{p} = \theta_L$ or $\tilde{p} = \theta_H$. In the former case, the product is immediately sold and the agent's payoff is $w_L = w_L(\theta_L)$. In the latter case, the customer offers the kickback $\beta\Delta\theta$, implying that the agent obtains at least $w_L + \beta\Delta\theta$. Hence, the agent offers the high price θ_H. He will reject the kickback and sell at the high price if the wage $w_H = w(\theta_H)$ is sufficiently large in comparison to the wage w_L to compensate the agent for his arguing costs, κ, and the forgone kickback. The corresponding no-collusion condition is

$$w_H - \kappa \geq w_L + \beta\left(\theta_H - \theta_L\right) \quad \Leftrightarrow \quad \Delta w \geq \beta\Delta\theta + \kappa \tag{NC}$$

with $\Delta w := w_H - w_L$. When condition (NC) does not hold, the agent accepts the kickback and sells at price θ_L. As the firm's contract $(w_\varnothing, w(p), \{\theta_L, \theta_H\})$ determines whether condition (NC) holds or not, the firm has to decide between preventing or allowing a possible

[17] The proof of Lemma 1 shows that any other deviation from $P = \{\theta_L, \theta_H\}$ cannot increase the firm's profit either. In practice, sales agents typically receive permission to grant discounts up to a certain amount, which would correspond to a price set of the form $P = [p_L, p_H]$. In our model, the price set $P = [\theta_L, \theta_H]$ would lead to the same outcomes as the price set $P = \{\theta_L, \theta_H\}$.

agent-customer collusion. We denote a delegation contract that prevents collusion as a contract of type D, and a delegation contract that allows for collusion is a contract of type D_{coll}.

We first characterize optimal incentive pay for contract type D, i.e., when the firm chooses a contract that satisfies the no-collusion constraint (NC). The firm's problem then reads as

$$\max_{w_\varnothing, w_L, w_H} g(a) \cdot (\theta_L + q\Delta\theta - w_L - q\Delta w) - (1 - g(a)) \cdot w_\varnothing \text{ subject to (NC),}$$

$$a \in \arg\max_{\hat{a}} g(\hat{a}) \cdot [w_L + q\Delta w - q\kappa] + (1 - g(\hat{a})) \cdot w_\varnothing - c(\hat{a}), \tag{7}$$

$$g(a) \cdot [w_L + q\Delta w - q\kappa] + (1 - g(a)) \cdot w_\varnothing - c(a) \geq 0, \tag{8}$$

$$w_\varnothing, w_L, w_H \geq 0. \tag{9}$$

Analogously to the case of centralization, the participation constraint (8) is implied by the limited-liability constraint (9). Furthermore, it is optimal to set $w_\varnothing = 0$. The incentive constraint (7) can be replaced by its corresponding first-order condition $W - q\kappa = c'(a)/g'(a)$, where $W = w_L + q\Delta w$ again denotes the expected wage. Altogether, using the function A as defined in Section 3.1, the firm's problem can be written as

$$\max_{w_L, w_H \geq 0} A\left(W - q\kappa\right) \cdot [\theta_L + q\Delta\theta - W] \quad \text{subject to (NC).} \tag{10}$$

The optimal solution to this problem may not be unique. The following proposition describes an optimal contract of type D.

Proposition 2 *Suppose the firm implements delegation without collusion, i.e., chooses contract type D. An optimal contract is then given by the feasible price set $P = \{\theta_L, \theta_H\}$, the wage $w_L^D = 0$ for selling at price θ_L, and the wage w_H^D for selling at price θ_H, where $w_H^D = \max\{W_D^*/q, \beta\Delta\theta + \kappa\}$ and W_D^* is implicitly described by*

$$\theta_L + q\Delta\theta = \frac{A(W_D^* - q\kappa)}{A'(W_D^* - q\kappa)} + W_D^*. \tag{11}$$

Proof. See Appendix. ■

The wages w_L and w_H serve two purposes, they provide effort incentives and prevent a collusion between the agent and a θ_H-customer. As w_L supports the first purpose but impedes the second one, whereas w_H supports both purposes, the firm optimally focuses compensation on w_H. When the no-collusion constraint (NC) is not binding (i.e., $W_D^*/q > \beta\Delta\theta + \kappa$), the two purposes are not in conflict. Collusion prevention then comes as a byproduct of high-powered effort incentives.[18] If, however, constraint (NC) is binding (i.e., $W_D^*/q \leq \beta\Delta\theta + \kappa$), in order to prevent collusion the firm implements incentive pay that makes the agent exert too much effort relative to the hypothetical situation where collusion is not an issue (i.e., if we could drop constraint (NC) from the firm's problem (10)).

Now suppose the firm chooses a contract of type D_{coll} so that the no-collusion constraint (NC) does not hold. The agent then offers the high price to a θ_H-customer, accepts the kickback $\beta\Delta\theta$ and sells at price θ_L. Hence, if a customer is found, the firm receives the price θ_L and pays the wage w_L to the agent independent of the customer's type. To ensure that (NC) is violated, the firm can choose $w_H = 0$ as optimal wage for selling at price θ_H.[19] In analogy to contract type D, we can set up and simplify the firm's problem. As the agent earns the wage w_L from selling to either customer type and additionally gets the kickback $\beta\Delta\theta$ when selling to a θ_H-customer, the firm solves:

$$\max_{w_L \geq 0} A\left(w_L + q\beta\Delta\theta\right) \cdot \left(\theta_L - w_L\right). \tag{12}$$

Proposition 3 describes the contract that solves problem (12).

Proposition 3 *Suppose the firm implements delegation with collusion, i.e., chooses contract type D_{coll}. An optimal contract is then given by $P = \{\theta_L, \theta_H\}$, $w_H^{D_{coll}} = 0$ as wage for selling at price θ_H, and wage $w_L^{D_{coll}}$ for selling at price θ_L, which is implicitly described by*

$$\theta_L = \frac{A\left(w_L^{D_{coll}} + q\beta\Delta\theta\right)}{A'\left(w_L^{D_{coll}} + q\beta\Delta\theta\right)} + w_L^{D_{coll}} \tag{13}$$

if $A\left(q\beta\Delta\theta\right)/A'\left(q\beta\Delta\theta\right) < \theta_L$, but $w_L^{D_{coll}} = 0$ otherwise.

Proof. See Appendix. ■

Contrary to contract type D, the agent never sells the product at a high price under contract type D_{coll}. However, allowing collusion has the advantage that the firm can use

[18] In this case, the optimal contract is not unique. The firm can choose all combinations of w_L and w_H that yield the optimal expected wage W_D^* and satisfy (NC).

[19] The optimal contract is not unique with respect to the specification of w_H. Alternatively, the firm can pay the agent a wage that is independent of the selling price or choose any other wage schedule that violates (NC).

implicit effort incentives via the expected kickback, $q\beta\Delta\theta$, which is paid by the customer and, hence, does not yield direct labor costs for the firm. Proposition 3 shows that explicit incentives via $w_L^{D_{coll}}$ and implicit incentives via $q\beta\Delta\theta$ are direct substitutes. If implicit incentives are small, the firm compensates the agent via explicit incentive pay. If implicit incentives $q\beta\Delta\theta$ increase, explicit incentives $w_L^{D_{coll}}$ will become smaller (see (13), holding θ_L constant). If implicit incentives reach a critical value such that $A(q\beta\Delta\theta)/A'(q\beta\Delta\theta) = \theta_L$, explicit incentive pay will be fully replaced by implicit incentives.

4 Centralization Versus Delegation

The previous section has identified four alternative contract types – two centralization contract types, C_L and C_H, as well as two delegation contract types, D and D_{coll} – and the corresponding conditionally optimal incentive pay as candidate solutions to the firm's contract design problem. In this section, we describe the firm's optimal contract design. Intuitively, centralization has the advantage that it prevents a possible collusion between the customer and the agent. The disadvantage of centralization is that the agent's knowledge on the customer's valuation will not be used. For example, when the firm has stipulated a price that turns out to be higher than the customer's valuation, as it can be the case under contract type C_H, the product cannot be sold. Moreover, anticipating that he might not be able to sell the product even if he finds a prospective customer, the agent's incentives to search for a customer decrease. By contrast, delegation allows the agent to adapt the price offer to the customer's valuation but may lead to collusion. The firm can prevent collusion by appropriately designing the agent's contract, but then the corresponding no-collusion constraint restricts the firm's contract space (contract type D). When the firm does not take precautions against collusion, it obtains a lower expected price for the product, but the prospect of earning a kickback can increase the agent's incentive to search for a customer (contract type D_{coll}).

Proposition 4 characterizes the optimal contract design depending on the probability that the customer has a high valuation for the product, q, and the agent's arguing costs, κ. The findings are illustrated in Figure 1.

Proposition 4 *(a) Suppose $q \leq \frac{\theta_L}{\theta_H}$. Contract type D is optimal if and only if $\kappa \leq (1-\beta)\,\Delta\theta$; otherwise contract type D_{coll} is optimal.*

(b) Suppose $q > \frac{\theta_L}{\theta_H}$. If $\kappa > (1-\beta)\Delta\theta$, only contract types D_{coll} and C_H can be optimal,

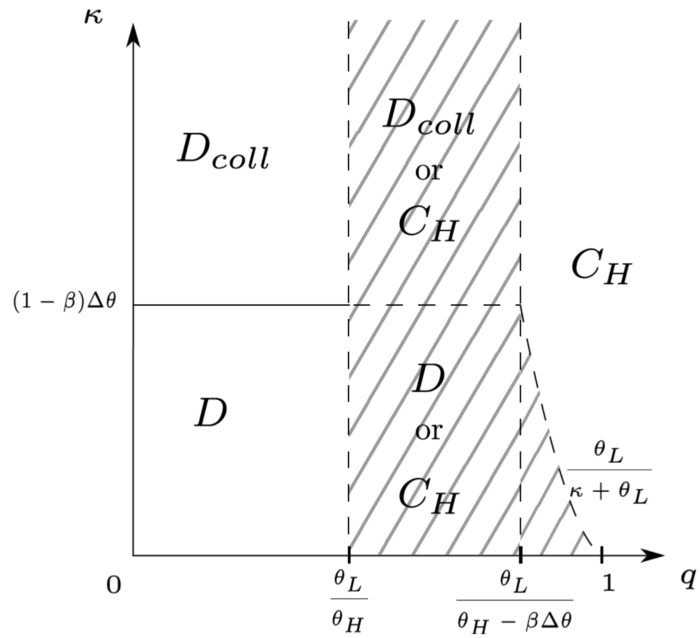

Figure 1: Optimal contract types

and contract type C_H dominates contract type D_{coll} if $q \geq \frac{\theta_L}{\theta_H - \beta\Delta\theta}$. If $\kappa \leq (1-\beta)\Delta\theta$, only contract types D and C_H can be optimal, and contract type C_H dominates contract type D if $q \geq \frac{\theta_L}{\kappa+\theta_L}$.

Proof. See Appendix. ■

According to Proposition 4, contract type C_L is never optimal. Intuitively, instead of choosing $P = \{\theta_L\}$ as the only feasible price, the firm is always better off by delegating pricing authority to the agent and allowing collusion, i.e., implementing contract type D_{coll}. The agent then also offers the price θ_L to either customer type but the firm benefits from the implicit effort incentives via the expected kickback. Thus, if it is sufficiently unlikely that the agent finds a θ_H-customer (i.e., $q \leq \theta_L/\theta_H$), delegation will dominate centralization from the firm's perspective. The firm then prefers the collusion-proof contract type D if and only if the θ_H-customer's additional willingness to pay, $\Delta\theta$, is sufficiently large relative to the forgone kickback, $\beta\Delta\theta$, and the agent's arguing costs, κ. This finding is intuitive as the firm has to compensate the agent for both $\beta\Delta\theta$ and κ to prevent collusion. Otherwise, the firm implements contract type D_{coll} in order to benefit from the implicit incentives via the prospective kickback. However, centralization with a high price can be optimal if it is sufficiently likely that the agent finds a θ_H-customer (i.e., $q > \theta_L/\theta_H$). Contract type C_H is certain to dominate delegation if θ_L is small because it is then rather unattractive to sell to a θ_L-customer.

Note that only contract type D exploits the agent's superior information about the customer's valuation of the product to apply price discrimination. Therefore, at first sight one might expect that D is particularly attractive for the firm if uncertainty about the customer type is high, i.e., if q takes intermediate values because using the agent's informational advantage is then most valuable.[20] For sufficiently low arguing costs κ, Proposition 4 confirms this conjecture, but in addition shows that contract type D is also favored by the firm when it is almost certain that a prospective customer has a low valuation of the product. The reason is that, as long as κ is not too large, replacing C_L with D is always worthwhile for the firm because the expected increase in sales, $q\Delta\theta$, dominates the costs of preventing collusion, $q\beta\Delta\theta + \kappa$. By contrast, when it is almost certain that a customer has a high valuation, replacing C_H with D is not beneficial because the expected increase in sales, $(1-q)\theta_L$, is too small relative to the costs of collusion prevention.

The proposition further shows that positive arguing costs κ are essential for the firm to benefit from collusion. When κ equals zero, the firm implements either contract type D or contract type C_H.[21] Contract type D then always dominates contract type D_{coll} because preventing collusion and obtaining a high price from a θ_H-customer is not very costly to the firm and hence more worthwhile than utilizing implicit incentives from kickbacks. This result also suggests that the firm allows collusion only when customers are sufficiently persistent when offering kickbacks.

Proposition 4 provides a sufficient condition for when the firm switches from delegation to centralization, namely, when the probability q that the agent encounters a θ_H-customer is sufficiently high. In order to make more precise predictions about when this kind of organizational change takes place and the subsequent adaption of the agent's incentive pay, we now consider parameterized functional forms for the probability of finding a customer, $g(a)$, and the agent's effort costs, $c(a)$. Specifically, we assume that $g(a) = a^\rho$ with effort $a \in [0, 1)$ and a parameter $\rho \in (0, 1]$. The higher ρ the more responsive the probability of

[20] In line with this argument, Lo et al. (2016) find some evidence that firms delegate more pricing authority when customer valuations are more variable.

[21] The condition $q \geq \frac{\theta_L}{\kappa+\theta_L}$ is sufficient but not necessary for C_H to be optimal. In the parameterized example below we show that C_H strictly dominates D for $\kappa = 0$ when q is sufficiently large and a high potential kickback $\beta\Delta\theta$ leads to strong incentives to collude (see Proposition 5, case (b)).

finding a customer is to effort.[22] Introducing the parameter ρ thus enables us to study how a varying effectiveness of the agent's search effort influences the optimal contract design. The agent's effort costs are $c(a) = \frac{a^2}{2c}$ with $c > 0$. We assume that the parameter c is sufficiently high so that the first-order condition of the agent's optimization problem always characterizes an optimal effort level that is strictly below one. Proposition 5 precisely describes when the firm switches from delegation to centralization, which according to Proposition 4 can happen only if $q > \frac{\theta_L}{\theta_H}$. The findings are illustrated in Figure 2.

Proposition 5 *Assume that $g(a) = a^\rho$ and $c(a) = \frac{a^2}{2c}$ with $a \in [0, 1)$, $\rho \in (0, 1]$, and $c > 0$.*

(a) Suppose that $\kappa > (1 - \beta)\Delta\theta$. There exists a threshold $\bar{q} \in \left(\frac{\theta_L}{\theta_H}, \frac{\theta_L}{\theta_H - \beta\Delta\theta}\right]$ such that contract type C_H dominates contract type D_{coll} if and only if $q \geq \bar{q}$. We have $\bar{q} = \frac{\theta_L}{\theta_H - \beta\Delta\theta}$ if and only if $\rho \geq 2\beta\frac{\Delta\theta}{\theta_L}$. The threshold $\bar{q}$ is always increasing in β and increasing in ρ as long as $\rho < 2\beta\frac{\Delta\theta}{\theta_L}$.

(b) Suppose that $\kappa \leq (1 - \beta)\Delta\theta$. There exists a threshold $\hat{q} \in \left(\frac{\theta_L}{\theta_H}, \frac{\theta_L}{\theta_L + \kappa}\right]$ such that contract type C_H dominates contract type D if and only if $q \geq \hat{q}$. We have $\hat{q} = \frac{\theta_L}{\theta_L + \kappa}$ if and only if $\rho \geq 2\beta\frac{\Delta\theta}{\theta_L}$. The threshold $\hat{q}$ is always decreasing in κ. It is decreasing in β and increasing in ρ as long as $\rho < 2\beta\frac{\Delta\theta}{\theta_L}$.

According to the comparative statics results of Proposition 5, the firm is more likely to implement delegation the more effective the agent's search effort, i.e., the higher the parameter ρ. Intuitively, as ρ increases, the agent's effort becomes more responsive to incentive pay,[23] which the firm thus wants to increase. When the agent obtains a higher reward for selling at a high price, he is less susceptible to kickbacks, meaning that contract type D becomes more attractive compared to C_H. A higher ρ also increases the comparative advantage of contract D_{coll} relative to C_H because, when the agent is more responsive to incentives, implicit incentives via kickbacks are even more useful from the firm's perspective. On the opposite scale, when ρ approaches zero, search effort does not play a role anymore because the agent will always be able to talk to a customer. The comparative advantage of contract type D_{coll} – implicit effort incentives through kickbacks – diminishes and D_{coll} becomes equivalent to centralization with a low price, represented by contract type C_L. Contract type D, however, still strictly dominates both centralization contract

[22] Note that $\rho = \frac{dg(a)}{da}\frac{a}{g(a)}$, i.e., ρ describes the effort elasticity of the probability function.

[23] One can show that the expected rent that the contract has to leave to the agent to induce a given effort level a is decreasing in ρ.

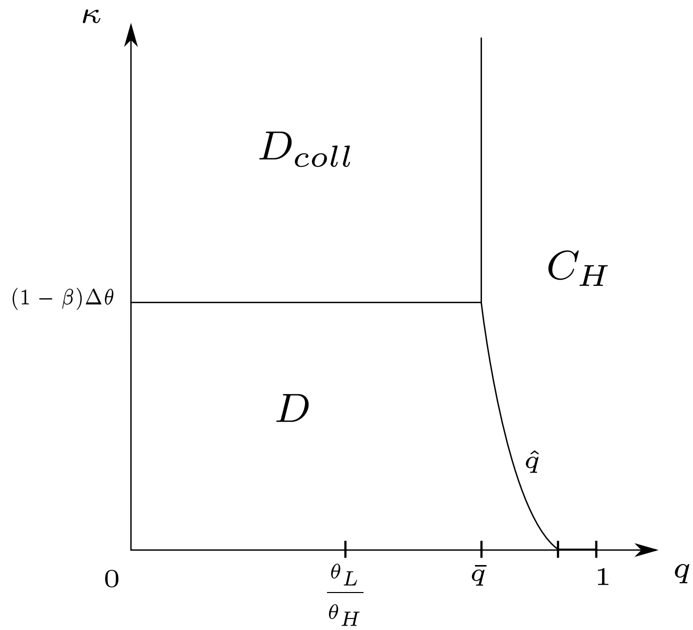

Figure 2: Optimal contract types for $c(a) = \frac{a^2}{2c}$ and $g(a) = a^\rho$

types C_L and C_H as long as q and κ are not too large.[24] The reason is that the comparative advantage of contract type D – utilizing the agent's superior information on the customer's willingness to pay – exists independently from the necessity to provide effort incentives.

At first sight, one might expect that delegation becomes less likely when the agent can appropriate a larger rent from collusion with customers. Proposition 5 shows that this intuition is not always true. When β and hence the size of a potential kickback increase, delegation becomes less likely when arguing costs are small (case (b)), but more likely when arguing costs are large (case (a)). In the former case, the firm wants to prevent collusion which now requires a larger wage for selling at a high price. In the latter case, the firm optimally allows collusion and thus indirectly benefits from higher kickbacks.

Finally, arguing costs κ affect the choice between delegation and centralization only as long as they are sufficiently small (case (b)). Contract type D then becomes less attractive relative to C_H as κ increases and thus higher incentive pay is needed to prevent collusion

[24]When $g(a) = 1$ for all a, under both contract types D_{coll} and C_L, the firm optimally sets wages equal to zero and earns profit θ_L. With respect to the comparison between contract type D and contract type C_H we obtain $\hat{q} = \frac{\theta_L}{\theta_H + \kappa - (1-\beta)\Delta\theta}$ (compare equation (21) for $\Gamma = 1$ in the proof of Proposition 5).

under D. By contrast, if κ is large (case (a)), it does not affect profits under the now relevant contract types D_{coll} and C_H because they either allow collusion or prevent it by means of centralization.

A central question in organizational design is how the delegation of decision rights and incentive pay interact. Should the firm accompany delegation with higher incentive pay, implying that delegation and monetary incentives are complements, or with lower incentive pay, meaning that delegation and monetary incentives are substitutes? Our model shows that, when a firm delegates pricing authority, the relationship between delegation and incentive pay is ambiguous, which is in line with the mixed empirical evidence on the interaction between the two instruments.[25]

Proposition 6 *Assume that $g(a) = a^\rho$ and $c(a) = \frac{a^2}{2c}$ with $a \in [0, 1)$, $\rho \in (0, 1]$, and $c > 0$.*

(a) Suppose that $\kappa > (1-\beta)\Delta\theta$ so that either contract type D_{coll} or C_H is optimal, and that the firm switches from C_H to D_{coll} due to an exogenous decrease of q. The firm then pays a smaller wage for a sale under D_{coll} than under C_H, i.e., $w_L^{D_{coll}} < w_H^*$. Moreover, the expected wage payment under D_{coll} is also smaller than under C_H.

(b) Suppose that $\kappa \leq (1-\beta)\Delta\theta$ so that either contract type D or C_H is optimal, and that the firm switches from C_H to D due to an exogenous decrease of q. When the no-collusion constraint (NC) *is binding under D, the firm pays a higher wage for selling at a high price under D than under C_H, i.e., $w_H^D > w_H^*$.*

Part (a) of Proposition 6 shows that delegation of pricing authority and incentive pay can be substitutes. Such a situation occurs when the delegation contract allows collusion and hence the firm at least partly relies on implicit effort incentives through kickbacks. Therefore, the agent's payment per sale is always lower under delegation than under centralization. The negative relationship between delegation and incentive pay persists if one looks at expected (or, in an empirical study, average) payments. This holds true even though, under delegation, the agent always earns a non-negative wage $w_L^{D_{coll}}$ when he has found a customer, whereas under centralization he obtains a positive wage w_H^* only if he has found a θ_H-customer. Moreover, under delegation prospective kickbacks contribute to

[25]While the first empirical studies find a complementary relation between decision-making authority and incentive pay (e.g., Nagar 2002; Wulf 2007; DeVaro and Kurtulus 2010), the recent papers by DeVaro and Prasad (2015) and Jia and van Veen-Dirks (2015) indicate that the interaction is not univocal.

the agent's search effort, so that the probability of finding a customer may be higher than under centralization.

Considering the wage payment for selling at a high price, part (b) of the proposition shows that delegation and incentives are complements when κ is small and the no-collusion constraint is binding.[26] The firm then needs high-powered incentives to prevent collusion, and hence the situation resembles a standard argument for a positive relationship between delegation and incentive pay: When the agent obtains more decision rights, high-powered incentives are necessary to preclude the agent from abusing his authority. With respect to average wage payments, however, the relationship is not clear cut. When the no-collusion constraint is not binding under D and hence optimal incentive pay is not unique, the relationship between delegation and incentive pay is also ambiguous. Because a potential collusion is then prevented as a byproduct of providing incentive pay, the standard argument for a positive relationship between delegation and incentive pay does not apply.

When studying the relationship between delegation and incentive pay, one can also compare a situation where the principal does not need to pay for performance because no moral hazard problem exists with a situation where performance pay is essential to elicit effort from the agent. Such a perspective is relevant for empirical predictions about the optimal variation of the degree of delegation between agents whose effort can be monitored (e.g., in-house sales agents contacting prospective customers via the phone) and agents whose effort is not observable (e.g., sales agents working in the field). When we take this perspective in our model, incentive pay and delegation are complements in the sense that the necessity to motivate effort (i.e., $\rho > 0$ instead of $\rho = 0$) leads to more delegation (i.e., $\bar{q}$ and $\hat{q}$ increase according to Proposition 5). Under contract type D, incentive pay that triggers the agent's effort also makes the agent less susceptible to collusion. Hence, preventing collusion becomes less costly for the firm relative to a situation where incentive pay is not necessary at all. With respect to contract type D_{coll}, moral hazard implies that the use of implicit incentives via kickbacks becomes beneficial from the firm's perspective, so that delegation with collusion can strictly dominate all other contract types. Consequently, delegation occurs more often in the presence of a moral hazard problem than under observable effort. This result is in line with an empirical study by Hong et al. (2015), who find that firms that adopt performance pay for exogenous reasons decentralize more decision-making authority to employees.

[26] Recall that, under contract type D, the optimal wage for selling at a low price is zero in this case.

5 Discussion

In this section, we consider several extensions of our basic model to discuss the robustness of the previous results.

5.1 Effort to Increase Customer Valuations

So far, we have assumed that the firm needs the agent to find a prospective customer. However, there are also situations where customers search for the product themselves and contact the firm. The agent might nevertheless be useful for the firm to increase a customer's valuation of the product and, hence, his willingness to pay. For example, during the sales talk the agent can exert effort to explain important features of the product and its various applications to the customer, which possibly increases the latter's valuation.[27] In this section, we analyze if and under which conditions the firm is still interested in price delegation to the agent in this alternative scenario.

Technically, we keep as many assumptions of our basic model as possible to make the results of this section comparable to our previous results. In particular, we only replace stage 3 of the game in Section 2 by the following assumptions: A customer contacts the sales agent and has the valuation $\theta = \theta_L > 0$ for the product. The agent chooses non-negative effort a at costs $c(a)$ to explain the useful features of the product to the customer. With probability $g(a)$ the agent is successful and the customer's valuation increases to $\theta = \theta_H > \theta_L$, but with probability $1 - g(a)$ the agent is not successful and the customer's valuation remains θ_L.[28] The technical properties of the functions $c(a)$ and $g(a)$ are identical to those of the respective functions of Section 2. We retain all the other assumptions of Section 2 as well as the previous notation.

Under this alternative type of moral hazard problem, price delegation in combination with collusion is always dominated by centralization, i.e., the firm does not want to make use of implicit effort incentives via expected kickbacks. To see this, consider contract type D_{coll} and let again p_L denote the lowest feasible price in P. If $p_L \leq \theta_L$, the agent always sells at price p_L, irrespective of whether he has increased the customer's valuation by $\Delta\theta$ or not. Hence, the firm cannot benefit from implicit incentives so that contract type

[27] Joseph (2001, p. 64) remarks that "Clearly, in many real-world situations, face-to-face communication may also involve negotiating and boosting willingness to pay." In Simester and Zhang (2014), the customer's willingness to pay is also affected by the sales agent's effort.

[28] In other words, we discard search effort and replace the exogenous probabilities q and $1 - q$ by the endogenous probabilities $g(a)$ and $1 - g(a)$, respectively.

D_{coll} is weakly dominated by a centralization contract that specifies $P = \{p_L\}$ and pays no wage. If $p_L > \theta_L$, the firm's expected profit reads as $A(w_L + \beta(\theta_H - p_L))(p_L - w_L)$, with w_L denoting the agent's wage for selling at price p_L.[29] Low values for p_L increase the agent's implicit incentives and, hence, the probability of increasing the customer's valuation, $A(w_L + \beta(\theta_H - p_L))$. However, low values of p_L also decrease the firm's net revenue in case of a sale, $p_L - w_L$. As the revenue effect always dominates the implicit-incentives effect, the firm optimally chooses the maximum value $p_L = \theta_H$ so that contract type D_{coll} is dominated by contract type C_H.[30]

A complete comparison of the different contract types leads to the following result:

Proposition 7 *Contract type D_{coll} is always (weakly) dominated.*

(a) There exists a threshold $\bar{\theta}_L \in (0, \theta_H)$ such that contract type C_H is optimal if and only if $\theta_L \leq \bar{\theta}_L$.

(b) If $\theta_L > \bar{\theta}_L$, only contract types D and C_L can be optimal. Contract type D will dominate contract type C_L if and only if $\kappa \leq (1 - \beta)\Delta\theta$.

Proposition 7 shows that the type of the moral hazard problem matters for the firm's optimal contract design. Contrary to Proposition 4, the firm no longer (strictly) prefers contract type D_{coll}. The contract type C_H will be optimal if θ_L is sufficiently small. Intuitively, C_H is the only contract type where the agent never sells at price θ_L. Therefore, if θ_L is large enough, either contract type D or C_L will be optimal. The firm prefers delegation via a collusion-proof contract to always selling at price θ_L if κ and β are sufficiently small. Small arguing costs, κ, and small values of the forgone kickback the agent has to be compensated for (i.e., a small β) imply that incentivizing the agent for selling at the high price θ_H is not too expensive for the firm. These two effects lead to exactly the same condition for the optimality of contract type D as in Proposition 4 (i.e., $\kappa \leq (1 - \beta)\Delta\theta$), given that θ_L is sufficiently large.

5.2 Imperfect Signal on Customer Valuations

In our basic model, the agent receives a perfect signal about the customer's valuation when he has found a customer. In this subsection, we sketch the implications of relaxing this

[29] Recall that the agent will never sell at price θ_H if collusion is possible.

[30] Technically, the revenue effect enters via the linear term $p_L - w_L$, whereas the implicit-incentives effect via the concave function $A(\cdot)$ and with weight $\beta \in (0, 1)$.

assumption. Agent and firm have the same prior distribution as before – a customer has a high (low) valuation with probability q $(1-q)$. However, the agent now learns a customer's valuation only with probability γ and has to rely on his prior information with probability $1-\gamma$. We assume that, under delegation, the firm again stipulates the feasible price set $P = \{\theta_L, \theta_H\}$. If the agent has not observed the customer's valuation and offers the price θ_H, a θ_L-customer will reject the offer and the game ends, whereas a θ_H-customer will offer the same kickback as in the basic model.

Obviously, the outcomes under centralization remain unchanged. However, outcomes will change in case of delegation. Under contract type D, the agent now chooses effort a to maximize

$$g(a) \cdot [\gamma \cdot (q(w_H - \kappa) + (1-q)w_L) + (1-\gamma) \cdot \max\{q(w_H - \kappa), w_L\}] - c(a).$$

The second term in squared brackets reflects that, with probability $1-\gamma$, the agent will offer a price without knowing the customer's type. When he offers a high price, his expected payoff is $q(w_H - \kappa)$. When he offers a low price, he obtains w_L for sure. An imperfect signal attenuates the agent's informational function under contract type D. Under a perfect signal, the agent always sells at a price that matches the customer's type. Under an imperfect signal, however, when not learning the customer's type, the agent either misses to sell to a θ_L-customer (if $q(w_H - \kappa) \geq w_L$) or sells at a too small price to a θ_H-customer (if $q(w_H - \kappa) < w_L$). Given an imperfect signal and contract type D_{coll}, the agent maximizes

$$g(a) \cdot [\gamma \cdot (w_L + q\beta\Delta\theta) + (1-\gamma) \cdot \max\{q(w_L + \beta\Delta\theta), w_L\}] - c(a).$$

The term $\max\{q(w_L + \beta\Delta\theta), w_L\}$ shows that the imperfect signal decreases the agent's incentives relative to the case of a perfect signal and, in addition, leads to an expected loss in sales if $q(w_L + \beta\Delta\theta) \geq w_L$.

All in all, an imperfect signal about the customer's valuation makes delegation of pricing authority less attractive to the firm, as the agent's ability to price discriminate between the different customer types is diminished. The relative advantage of each type of delegation contract, however, remains qualitatively the same.

5.3 More Than Two Customer Types

We have assumed so far that a customer can have two possible types. For example, such a situation occurs when the good has a special feature that is useful only to a share of the prospective customers, who assign the same value to this feature. When a customer can have more than two possible valuations for the good, the key trade-offs of our basic model remain in place. To explain this point, we first sketch the firm's optimal contract design for the case of three possible types, i.e., $\theta \in \{\theta_L, \theta_M, \theta_H\}$ with $0 < \theta_L < \theta_M < \theta_H$. As in our basic model, the firm always implements a collusion-proof contract when the agent has no arguing costs ($\kappa = 0$). The firm then either chooses centralization and stipulates the fixed price θ_H, or implements a delegation contract with $P = \{\theta_M, \theta_H\}$ or $P = \{\theta_L, \theta_M, \theta_H\}$. Intuitively, starting from the highest reasonable price θ_H, the firm needs to decide whether the agent should be granted the authority to offer a price discount and, if so, to what extent. It is never optimal to exclude θ_H from the feasible price set because the ensuing advantage, namely lower expected costs for preventing collusion with a θ_H-customer, is dominated by the associated decrease of the expected sales price. Analogously, it is never optimal to exclude the price θ_M when the firm allows to charge the price θ_L. These results may no longer hold true when the costs of collusion prevention increase due to positive arguing costs ($\kappa > 0$). The above contracts can then be dominated by centralization contracts with fixed prices θ_L or θ_M. However, these centralization contracts are in turn dominated by delegation contracts that allow for collusion. These contracts include θ_L or θ_M in the feasible price set but also allow for higher prices so that the agent can earn kickbacks, which enhances his effort incentives. Hence, the main trade-offs of our basic model remain in place, but the set of contracts that can be optimal depending on the model parameters increases. The same rationale applies to a general but finite number of customer types.

When the customer's type is continuous such that $\theta \in [\theta_L, \theta_H]$ and the agent has no arguing costs, the firm again always prevents collusion. However, in contrast to the discrete case, it can be shown that the agent will always obtain some pricing authority.[31] The firm then stipulates a feasible interval of prices $[p_L, \theta_H]$ with $\theta_L \leq p_L < \theta_H$. If $\kappa > 0$ and hence preventing collusion becomes more costly, the firm may prefer to centralize pricing authority and fix a price from the interval $[\theta_L, \theta_H]$. However, the firm is then even better off by allowing the agent to offer higher prices and allowing collusion. Hence, again, the key trade-offs from our basic model remain existent.

[31] The proof is available from the authors upon request.

5.4 Collusion Detection

In our basic model, we have not considered the possibility that the firm can detect and punish collusion. Implementing a mechanism to discover collusion is worthwhile for the firm only if the associated monitoring costs are smaller than the expected wage savings relative to an optimal contract under contract type D in the basic model. In particular, when the no-collusion constraint (NC) is not binding, implying that the firm achieves collusion prevention as a byproduct of incentive provision, a collusion detection mechanism does not increase firm profits. If constraint (NC) is binding and the costs of implementing a collusion detection mechanism are sufficiently small, we can easily extent the model to include collusion detection. To see this, assume that the firm discovers collusion with probability $\tau \in (0,1)$ and, in case collusion is revealed, the agent does not obtain a wage and cannot keep the kickback.[32] Hence, the no-collusion constraint under contract type D changes to $w_H - \kappa \geq (1-\tau)(w_L + \beta\Delta\theta)$. The possibility to reveal collusion makes contract type D more attractive relative to the other contract types because it lowers the firm's costs of collusion prevention. Consequently, contract type D will be implemented the more often the more effective the collusion detection mechanism, i.e., the higher τ. Contract type D_{coll} nevertheless remains optimal for sufficiently large values of κ, while contract type C_H continues to be optimal for sufficiently high values of q.

5.5 Risk Aversion and Positive Outside Options

Standard models on moral hazard typically suggest two alternative frictions to exclude trivial solutions to the contracting problem in which the firm always implements first-best efforts (Laffont and Martimort, 2002). First, the agent is assumed to be risk neutral but protected by limited liability. Second, the agent is assumed to be risk averse but has unlimited liability. In this paper, we decided to make use of the first contractual friction because it facilitates the analysis and sales agents can be expected to be less risk averse than other employees. As both kinds of contractual frictions make the creation of incentives costly for the firm, they should lead to qualitatively similar outcomes. However, under risk aversion and unlimited liability, the firm would favor contract type D_{coll} for a different reason than in our model. Our results have shown that D_{coll} yields implicit incentives for the agent that are paid by high-valuation customers. The firm benefits from these implicit incentives because they can replace explicit incentive pay and hence lower the rent

[32] Limited liability of the agent prevents the firm from imposing fines.

that the firm has to leave to the agent. If the agent is not protected by limited liability, rent payments are not an issue. However, the agent's participation constraint is binding under the optimal contract and implicit incentives from kickbacks may relax this binding constraint. In other words, the firm has to offer lower average wage payments to make the agent sign the contract.

A similar effect can arise if the agent is protected by limited liability but has a sufficiently large reservation value. In that situation, the firm may be forced to offer the agent a positive lump-sum payment in addition to the incentive pay to ensure that he signs the contract. The optimal lump-sum payment makes the participation constraint just bind. Then the contract type D_{coll} exhibits the same beneficial participation argument as under risk aversion and unlimited liability, as the implicit incentives lead to a relaxation of the binding constraint. This participation argument for the use of contract type D_{coll} might even make this contract type optimal for the firm when it is clearly dominated for incentive reasons as in Section 5.1.

6 Conclusion

Industries employing sales agents that work in the field face a severe moral hazard problem, as the agents have large discretion over their effort for finding a prospective customer. At first sight, one might expect that these firms should not further increase agents' discretion by delegating pricing authority to them. However, our model yields a different prediction. The sales agents already need high-powered incentives as a measure against moral hazard. If they receive pricing authority, the firm can specify a large extra bonus for selling at a high price. These high-powered incentives, which work against moral hazard, then also prevent a possible collusion between customers offering a kickback and agents in turn granting an unnecessary price discount. In other words, high-powered incentives that prevent sales agents from shirking can be used in a complementary way to prevent collusion. Thus, our model predicts that industries employing sales agents in the field should use higher-powered incentives and more delegation of pricing authority compared to industries employing sales agents that can be better monitored (e.g., in-house sales agents). Moreover, the more effective the sales agents' prospecting effort, the more likely should a firm be to delegate.

These predictions even continue to hold when it becomes too expensive for a firm to prevent collusion. In this case, expected rents paid by customers to an agent with price discretion help mitigate moral hazard. Hence, a firm again finds it optimal to delegate pricing authority when moral hazard problems are severe, or the agent's effort is very productive. Finally, our theoretical analysis points out that, in the presence of collusion, empirical predictions regarding the optimal interaction between the delegation of pricing authority and incentive pay need to be treated cautiously. We have shown that the optimal interplay between the two instruments crucially depends on whether the firm implements collusion-proof contracts or not. Such information will in general not be readily available for an empirical analysis.

7 Appendix

Proof of Proposition 2. Neglect for a moment the no-collusion constraint (NC). As the firm's objective function – described by (10) – is strictly concave, the optimal expected wage W_D^* is given by the first-order condition

$$\theta_L + q\Delta\theta = \frac{A\left(W_D^* - q\kappa\right)}{A'\left(W_D^* - q\kappa\right)} + W_D^*.$$

Thus, if the Δw that is implicitly described by

$$\theta_L + q\Delta\theta = \frac{A\left(q\Delta w - q\kappa\right)}{A'\left(q\Delta w - q\kappa\right)} + q\Delta w$$

also satisfies (NC), optimal wages that solve the firm's constrained maximization problem will be $w_L^D = 0$ and $w_H^D = \Delta w$. Otherwise, the no-collusion constraint (NC) is binding, leading to the (uniquely) optimal wages $w_L^D = 0$ and $w_H^D = \beta\Delta\theta + \kappa$. ■

Proof of Proposition 3. As the firm's objective function (12) is strictly concave, the optimal wage $w_L^{D_{coll}}$ will be strictly positive and described by the first-order condition (13) if the objective function's derivative with respect to w_L is positive at $w_L = 0$, i.e., if $A\left(q\beta\Delta\theta\right)/A'\left(q\beta\Delta\theta\right) < \theta_L$. Otherwise, there is a corner solution at $w_L^{D_{coll}} = 0$. ■

Proof of Proposition 4. In order to facilitate the comparison of the four contract types, we first reformulate the firm's corresponding optimization problems in terms of choosing the agent's expected net payoff Y given that the agent has already chosen effort and found a customer. Accordingly, Y includes the expected wage payment by the firm and, in case of delegation, any expected costs (i.e., $q\kappa$ under contract type D) or benefits (i.e., the kickback $q\beta\Delta\theta$ under contract type D_{coll}) from interacting with the customer.

Under *centralization*, we have $Y = W$ and analogous to problem (4) the firm solves a problem of the form

$$\max_{Y \geq 0} A(Y)(\Theta - Y),$$

where $\Theta = \Theta_{C_H} := q\theta_H$ under contract type C_H and $\Theta = \Theta_{C_L} := \theta_L$ under contract type C_L. Contract type C_H will dominate contract type C_L iff $\Theta_{C_H} \geq \Theta_{C_L}$ or $q\theta_H \geq \theta_L$, as we already know from Proposition 1.

Under *delegation*, problems (10) and (12) can be written as

$$\max_{Y} A(Y)(\Theta - Y) \quad \text{subject to } Y \geq q\beta\Delta\theta. \tag{14}$$

Under contract type D, we already set $w_L = 0$ (which is optimal according to Proposition 2) so that $Y = qw_H - q\kappa$ and $\Theta = \Theta_D := \theta_L + q\Delta\theta - q\kappa$. The restriction in (14) reflects the no-collusion constraint (NC). Under contract type D_{coll}, setting $w_H = 0$, we have $Y = w_L + q\beta\Delta\theta$ and $\Theta = \Theta_{D_{coll}} := \theta_L + q\beta\Delta\theta$. Here, the restriction in (14) reflects the limited liability constraint $w_L \geq 0$. From this specification of the firm's optimization problem, it follows that contract type D dominates contract type D_{coll} iff

$$\Theta_D \geq \Theta_{D_{coll}} \Leftrightarrow \kappa \leq (1-\beta)\Delta\theta.$$

The following result shows that contract type C_L can never be optimal:

Lemma A1 *Contract type D_{coll} dominates contract type C_L.*

Proof of Lemma A1. First suppose the restriction $Y \geq q\beta\Delta\theta$ is not binding under D_{coll}. Then, contract type D_{coll} dominates C_L because $\Theta_{D_{coll}} > \Theta_{C_L}$. Now suppose the restriction is binding, i.e., $Y^* = q\beta\Delta\theta$. From Proposition 3 we know that this is the case if $\frac{A(q\beta\Delta\theta)}{A'(q\beta\Delta\theta)} \geq \theta_L$. By Proposition 1, under contract type C_L we have $\theta_L = \frac{A(w_L^*)}{A'(w_L^*)} + w_L^*$. Thus, because $\frac{A(\cdot)}{A'(\cdot)}$ is an increasing function, we obtain $q\beta\Delta\theta > w_L^*$. Therefore, we must have that $A(q\beta\Delta\theta) \cdot \theta_L > A(w_L^*) \cdot (\theta_L - w_L^*)$, i.e., expected firm profits are larger under D_{coll} than under C_L. ■

Now, Proposition 4 can be proved as follows. First, consider $q\theta_H \leq \theta_L$. In this case, C_L dominates C_H but the former contract type can never be optimal (Lemma A1). Hence, only D or D_{coll} can be optimal. From above we know that D will dominate D_{coll} iff $\kappa \leq (1-\beta)\Delta\theta$. Second, consider $q\theta_H > \theta_L$ and $\kappa > (1-\beta)\Delta\theta$. By the previous results, either D_{coll} or C_H is optimal. A sufficient condition for C_H dominating D_{coll} is that

$$\Theta_{C_H} \geq \Theta_{D_{coll}} \Leftrightarrow q \geq \frac{\theta_L}{\theta_H - \beta\Delta\theta}.$$

Finally, consider $q\theta_H > \theta_L$ and $\kappa \leq (1-\beta)\Delta\theta$. In that case, either D or C_H is optimal. A sufficient condition for C_H dominating D is that

$$\Theta_{C_H} \geq \Theta_D \Leftrightarrow q \geq \frac{\theta_L}{\kappa + \theta_L}.$$

■

Proof of Proposition 5. The proof builds on the proof of Proposition 4. From Proposition 4 it already follows that $\bar{q}, \hat{q} \geq \frac{\theta_L}{\theta_H}$.

(a) In this case, contract type D_{coll} dominates contract type D. If $q = \frac{\theta_L}{\theta_H}$, then the contract types C_L and C_H lead to the same profit (compare problem (4)). Hence, because D_{coll} strictly dominates C_L (compare Lemma A1), D_{coll} also strictly dominates C_H if $q = \frac{\theta_L}{\theta_H}$. By continuity, D_{coll} also strictly dominates C_H for some values $q > \frac{\theta_L}{\theta_H}$, implying that $\bar{q} > \frac{\theta_L}{\theta_H}$.

According to Proposition 4, a sufficient condition for C_H dominating D_{coll} is that $q \geq \frac{\theta_L}{\theta_H - \beta\Delta\theta} =: q_{max}$. Define $\tilde{q}$ such that $\frac{A(\tilde{q}\beta\Delta\theta)}{A'(\tilde{q}\beta\Delta\theta)} = \theta_L$. According to Proposition 3 and because $A(\cdot)/A'(\cdot)$ is an increasing function, the restriction $Y \geq q\beta\Delta\theta$ is binding under D_{coll} iff $q \geq \tilde{q}$. If $q_{max} \leq \tilde{q}$, then D_{coll} dominates C_H for all $q < q_{max}$ and hence we obtain for the threshold $\bar{q}$ that $\bar{q} = q_{max}$.

If $q_{max} > \tilde{q}$, then C_H dominates D_{coll} for some values $q < q_{max}$ but not for $q = \tilde{q}$. Hence, the optimal profit functions under C_H and D_{coll} have at least one intersection, which is the candidate for the threshold $\bar{q}$. We now show that, for the given functions $g(a) = a^\rho$ and $c(a) = \frac{a^2}{2c}$, there is exactly one intersection q_s. We obtain $\frac{c'(a)}{g'(a)} = \frac{1}{c\rho}a^{2-\rho}$ and for the corresponding inverse function $e(Y) = (c\rho Y)^{\frac{1}{2-\rho}}$, where Y denotes the agent's expected net payoff as defined in the proof of Proposition 4. It follows that $A(Y) = [e(Y)]^\rho = (c\rho Y)^{\frac{\rho}{2-\rho}}$ and $\frac{A(Y)}{A'(Y)} = \frac{2-\rho}{\rho}Y$. Hence, we have $\tilde{q} = \frac{\rho\theta_L}{(2-\rho)\beta\Delta\theta}$, and $q_{max} > \tilde{q}$ is equivalent to $\frac{2}{\rho}\frac{\Delta\theta}{\theta_H}\beta > 1$. Under D_{coll}, the optimal profit function for $q \geq \tilde{q}$ is

$$\Pi^{D_{coll}} = A(q\beta\Delta\theta)\theta_L = (c\rho q\beta\Delta\theta)^{\frac{\rho}{2-\rho}}\theta_L. \tag{15}$$

Under C_H, from Proposition 1 it follows that $w_H^* = \frac{\rho}{2}\theta_H$. Consequently, the optimal profit function is

$$\Pi^{C_H} = A\left(q\frac{\rho}{2}\theta_H\right)\left(q\theta_H - q\frac{\rho}{2}\theta_H\right) = \left(cq\frac{\rho^2}{2}\theta_H\right)^{\frac{\rho}{2-\rho}} q\theta_H\frac{2-\rho}{2}. \tag{16}$$

We obtain that $\Pi^{C_H} \geq \Pi^{D_{coll}}$ iff

$$q \geq \frac{\theta_L}{\theta_H}\frac{2}{2-\rho}\left(\frac{2\beta\Delta\theta}{\rho\theta_H}\right)^{\frac{\rho}{2-\rho}} =: q_s = \bar{q}. \tag{17}$$

Clearly, $\frac{\partial q_s}{\partial\beta} > 0$. Using a mathematical computation program, one can verify that $\frac{\partial q_s}{\partial\rho} > 0$ in the relevant parameter range (i.e., $\frac{2}{\rho}\frac{\Delta\theta}{\theta_H}\beta > 1$).

(b) In this case, D dominates D_{coll} and, hence, also C_H for some values $q > \frac{\theta_L}{\theta_H}$. Following Proposition 4, a sufficient condition for C_H dominating D is that $q \geq \frac{\theta_L}{\theta_L+\kappa} =: q^D_{max}$.

When we neglect the restriction $Y \geq q\beta\Delta\theta$ for D in (14), the optimal Y is given by $\Theta_D = \frac{A(Y)}{A'(Y)} + Y$, implying $Y = \frac{\rho}{2}\Theta_D = \frac{\rho}{2}(\theta_L + q(\Delta\theta - \kappa))$. Hence, the restriction is binding iff

$$\frac{\rho}{2}(\theta_L + q(\Delta\theta-\kappa)) \leq q\beta\Delta\theta \quad \Leftrightarrow \quad \theta_L \leq q\left[\left(\frac{2\beta}{\rho}-1\right)\Delta\theta + \kappa\right]. \tag{18}$$

If the right-hand side of the last inequality is negative or zero, the restriction is never binding and, hence, $\hat{q} = q^D_{max}$. Now consider the case where the right-hand side is positive. Then, the restriction is binding iff $q \geq \tilde{q}^D$ with

$$\tilde{q}^D := \frac{\theta_L}{\left(\frac{2\beta}{\rho}-1\right)\Delta\theta + \kappa}. \tag{19}$$

If $q^D_{max} \leq \tilde{q}^D$, we again have $\hat{q} = q^D_{max}$. Now suppose that $q^D_{max} > \tilde{q}^D$, which is equivalent to $\frac{2}{\rho}\frac{\Delta\theta}{\theta_H}\beta > 1$. In this case, C_H dominates D for some $q < q^D_{max}$ but not for $q = \tilde{q}^D$. Hence, the optimal profit functions intersect at least once. We now show that there is only one intersection. Under D, the optimal profit function for $q \geq \tilde{q}^D$ is

$$\Pi^D = A(q\beta\Delta\theta)(\theta_L + q(\Delta\theta-\kappa) - q\beta\Delta\theta) = (c\rho q\beta\Delta\theta)^{\frac{\rho}{2-\rho}}(\theta_L + q((1-\beta)\Delta\theta - \kappa)). \tag{20}$$

It follows that $\Pi^{C_H} \geq \Pi^D$ iff[33]

$$q \geq \frac{\Gamma\theta_L}{\frac{2-\rho}{2}\theta_H - \Gamma[(1-\beta)\Delta\theta - \kappa)]} =: q_s^D = \hat{q}, \text{ with } \Gamma := \left(\frac{2\beta\Delta\theta}{\rho\theta_H}\right)^{\frac{\rho}{2-\rho}}. \qquad (21)$$

Clearly, $\frac{\partial q_s^D}{\partial \kappa} < 0$. Using a mathematical computation program, it can be verified that $\frac{\partial q_s^D}{\partial \rho} > 0$ and $\frac{\partial q_s^D}{\partial \beta} < 0$ in the relevant parameter range (i.e., $\frac{2}{\rho}\frac{\Delta\theta}{\theta_H}\beta > 1$). ■

Proof of Proposition 6. (a) According to Proposition 1, the wage payment for a sale under C_H is $w_H^* = \frac{\rho}{2}\theta_H$. According to Proposition 3, the optimal wage for a sale at a low price, $w_L^{D_{coll}}$, is either equal to zero or equal to

$$\frac{\rho}{2}\left(\theta_L - \frac{2-\rho}{\rho}q\beta\Delta\theta\right).$$

Both values are strictly smaller than $w_H^* = \frac{\rho}{2}\theta_H$ for all q. The proof of Proposition 5 has shown that, given an expected net payoff Y for the agent, the probability of finding a customer is $A(Y) = (c\rho Y)^{\frac{\rho}{2-\rho}}$. The expected wage payment under C_H is given by $A(qw_H^*)qw_H^* = A\left(q\frac{\rho}{2}\theta_H\right)q\frac{\rho}{2}\theta_H$. The expected wage under D_{coll} is given by $A(w_L^{D_{coll}} + q\beta\Delta\theta)w_L^{D_{coll}}$. For $w_L^{D_{coll}} > 0$, we obtain

$$A(w_L^{D_{coll}} + q\beta\Delta\theta)w_L^{D_{coll}} = A\left(\frac{\rho}{2}\theta_L + \left(1 - \frac{\rho}{2}\frac{2-\rho}{\rho}\right)q\beta\Delta\theta\right)w_L^{D_{coll}} = A\left(\frac{\rho}{2}\left[\theta_L + q\beta\Delta\theta\right]\right)w_L^{D_{coll}},$$

which can be shown to be decreasing in q. Hence, $A\left(\frac{\rho}{2}\theta_L\right)\frac{\rho}{2}\theta_L$ is an upper bound for the expected wage payment under D_{coll}. By Proposition 4, C_H can dominate D_{coll} only for those values of q that satisfy $q\theta_H > \theta_L$. Hence, because $A(\cdot)$ is an increasing function, we obtain that $A\left(q\frac{\rho}{2}\theta_H\right)q\frac{\rho}{2}\theta_H > A\left(\frac{\rho}{2}\theta_L\right)\frac{\rho}{2}\theta_L$, i.e., the expected wage payment under C_H is larger than under D_{coll}.

(b) According to Proposition 2, an optimal incentive scheme under D is given by $w_L^D = 0$ and

$$\begin{aligned} w_H^D &= \max\left\{\frac{1}{q}\frac{\rho}{2}\left[\theta_L + q\Delta\theta + \frac{2-\rho}{\rho}q\kappa\right], \beta\Delta\theta + \kappa\right\} \\ &= \max\left\{\frac{\rho}{2}\left[\frac{\theta_L}{q} - \theta_L + \theta_H + \frac{2-\rho}{\rho q}q\kappa\right], \beta\Delta\theta + \kappa\right\} > \frac{\rho}{2}\theta_H = w_H^*. \end{aligned}$$

[33] Recall that there is at least one intersection, which ensures that the denominator of the following threshold is positive.

In case the second term in curly brackets is larger than the first term, the no-collusion constraint (NC) is binding and (w_L^D, w_H^D) is the uniquely optimal incentive scheme under D. ■

Proof of Proposition 7. In analogy to Section 3, the nonnegative wages always imply the agent's participation constraint, and the firm optimally chooses $w_\varnothing = 0$ under each contract type C_L, C_H, D_{coll}, and D.

(1) In case of centralization, again the firm chooses either $P = \{\theta_L\}$ or $P = \{\theta_H\}$ as feasible prices, because any other singleton would leave a positive rent to the customer or prevent a sale. If $P = \{\theta_L\}$, the agent will sell to the customer at price θ_L for sure and does not exert positive effort. Consequently, contract type C_L leads to the optimal profit $\Pi^{C_L} = \theta_L$.

(2) Now consider contract type C_H where $P = \{\theta_H\}$. For the given wage $w_H := w(\theta_H)$, the agent maximizes $g(a) \cdot w_H - c(a)$, yielding $a = e(w_H)$ as incentive constraint. Thus, w_H is chosen by the firm to maximize $A(w_H)(\theta_H - w_H)$. As this objective function is strictly concave, the optimal wage is positive, leading to the optimal profit $\Pi^{C_H} > 0$.

(3) Next, we consider the case of delegating and allowing collusion. Suppose the lowest feasible price satisfies $p_L \leq \theta_L$. In that case, the firm does not benefit from implicit and explicit incentives, as the agent always sells at price p_L, irrespective of the customer's valuation of the product. Thus, delegation in combination with collusion is weakly dominated by a centralization contract that specifies $P = \{p_L\}$ and pays no wage. Suppose the firm chooses $p_L > \theta_L$. Then the agent maximizes $g(a) \cdot (w_L + \beta \cdot (p_H - p_L)) - c(a)$ with $w_L := w(p_L)$, yielding $a = e(w_L + \beta(p_H - p_L))$ as incentive constraint. The firm's profit can be written as

$$A(w_L + \beta(p_H - p_L))(p_L - w_L).$$

Setting $p_H = \theta_H$ is optimal for the firm to maximize the agent's implicit incentives from earning a kickback. As $A(w_L + \beta(\theta_H - p_L))(p_L - w_L)$ is strictly concave in w_L, the optimal wage is either positive – being denoted by $\hat{w}_L > 0$ – and implicitly described by the first-order condition

$$A(\hat{w}_L + \beta(\theta_H - p_L)) = A'(\hat{w}_L + \beta(\theta_H - p_L))(p_L - \hat{w}_L), \tag{22}$$

or it is zero (i.e., $\hat{w}_L = 0$) given that $\frac{\partial}{\partial w_L} A(w_L + \beta(\theta_H - p_L))(p_L - w_L) \leq 0$ at $w_L = 0$, i.e.,

$$A'(\beta(\theta_H - p_L))\, p_L \le A(\beta(\theta_H - p_L)). \tag{23}$$

In case of an interior solution, $\hat{w}_L > 0$, we can apply the Envelope Theorem to analyze how an increase of p_L influences the optimal profit $\hat{\Pi} = A(\hat{w}_L + \beta(\theta_H - p_L))(p_L - \hat{w}_L)$:

$$\begin{aligned}\frac{\partial \hat{\Pi}}{\partial p_L} &= -\beta A'(\hat{w}_L + \beta(\theta_H - p_L))(p_L - \hat{w}_L) + A(\hat{w}_L + \beta(\theta_H - p_L)) \\ &\overset{(22)}{=} (1-\beta)\, A(\hat{w}_L + \beta(\theta_H - p_L)) > 0.\end{aligned}$$

Thus, in case of an interior solution, the firm prefers $p_L = \theta_H$ so that C_H dominates delegation in combination with collusion. The same is also true in case of a corner solution, $\hat{w}_L = 0$. For the respective optimal profit $\hat{\Pi} = A(\beta(\theta_H - p_L))\, p_L$ we obtain

$$\frac{\partial \hat{\Pi}}{\partial p_L} = -\beta A'(\beta(\theta_H - p_L))\, p_L + A(\beta(\theta_H - p_L)) \overset{(23)}{>} (1-\beta)\, A(\beta(\theta_H - p_L)) > 0.$$

To sum up, contract type D_{coll} is always (weakly) dominated by centralization.

(4) Under a contract of type D, the firm chooses $P = \{\theta_L, \theta_H\}$, a zero wage for selling at price θ_L, and a positive wage $w_H = w(\theta_H)$ for selling at price θ_H.[34] The agent maximizes $g(a) \cdot (w_H - \kappa) - c(a)$, yielding $a = e(w_H - \kappa)$ as incentive constraint. The firm chooses w_H to maximize profit

$$A(w_H - \kappa)(\theta_H - w_H) + (1 - A(w_H - \kappa))\, \theta_L = \theta_L + A(w_H - \kappa)(\Delta\theta - w_H)$$

subject to the no-collusion constraint $w_H \ge \beta\Delta\theta + \kappa$. If, at the optimum, the no-collusion constraint is non-binding, the corresponding wage $w_H^D > \beta\Delta\theta + \kappa$ is implicitly described by the first-order condition

$$A'\left(w_H^D - \kappa\right)\left(\Delta\theta - w_H^D\right) = A\left(w_H^D - \kappa\right), \tag{24}$$

as the firm's profit function is strictly concave. In case of a binding constraint, the optimal

[34] $P = \{\theta_L, \theta_H\}$ follows in analogy to Case (I) from the Supplementary Material. In particular, a price $p_H < \theta_H$ would lower the forgone kickback the agent has to be compensated for but also the firm's revenue from selling to a θ_H-customer, and the second effect always dominates the first. A wage $w(\theta_L) > 0$ would only decrease incentives and increase labor costs. Furthermore, $w(\theta_L) = 0$ relaxes the no-collusion constraint below.

wage $w_H^D = \beta\Delta\theta + \kappa$ lies to the right of the wage that maximizes $\theta_L + A(w_H - \kappa)(\Delta\theta - w_H)$ and, thus, satisfies

$$A'(\beta\Delta\theta)((1-\beta)\Delta\theta - \kappa) < A(\beta\Delta\theta). \tag{25}$$

Comparison of firm profit under D with Π^{C_L} yields that

$$\theta_L + A(w_H - \kappa)(\Delta\theta - w_H) \geq \Pi^{C_L} \Leftrightarrow w_H \leq \Delta\theta.$$

Contract type D will dominate contract type C_L if and only if there exist values of w_H that satisfy $w_H \leq \Delta\theta$ together with $w_H \geq \beta\Delta\theta + \kappa$, i.e., iff $\kappa \leq (1-\beta)\Delta\theta$.

Comparison of $\Pi^{C_L} = \theta_L$ and Π^{C_H} shows that $\Pi^{C_L} < \Pi^{C_H}$ for $\theta_L \to 0$, and $\Pi^{C_L} > \Pi^{C_H}$ for $\theta_L \to \theta_H$, because Π^{C_H} is independent of θ_L and satisfies $\Pi^{C_H} \in (0, \theta_H)$. Thus, as Π^{C_L} is monotonically increasing in θ_L, there exists a threshold $\tilde{\theta}_L \in (0, \theta_H)$ such that $\Pi^{C_L} < (>)\Pi^{C_H}$ iff $\theta_L < (>)\tilde{\theta}_L$.

Finally, we have to compare the optimal profit under contract type D, Π^D, with the optimal profit under contract type C_H, Π^{C_H}. In case of contract type D and a non-binding no-collusion constraint, by applying the Envelope Theorem we have

$$\frac{\partial \Pi^D}{\partial \theta_L} = 1 - A\left(w_H^D - \kappa\right) > 0$$

with $A\left(w_H^D - \kappa\right) < 1$ describing a probability. In addition, if the no-collusion constraint is binding, we obtain

$$\begin{aligned}\frac{\partial \Pi^D}{\partial \theta_L} &= 1 - \beta A'(\beta\Delta\theta)((1-\beta)\Delta\theta - \kappa) - (1-\beta)A(\beta\Delta\theta) \\ &\overset{(25)}{>} 1 - \beta A(\beta\Delta\theta) - (1-\beta)A(\beta\Delta\theta) = 1 - A(\beta\Delta\theta) > 0.\end{aligned}$$

Therefore, irrespective of whether the no-collusion constraint is binding or not at the optimum, the optimal profit Π^D is strictly increasing in θ_L. Given $w_H^D > \beta\Delta\theta + \kappa$, for $\theta_L \to 0$ we obtain

$$\Pi^D = A\left(w_H^D - \kappa\right)\left(\theta_H - w_H^D\right) < A\left(w_H^D\right)\left(\theta_H - w_H^D\right) \leq \Pi^{C_H}.$$

Given $w_H^D = \beta\Delta\theta + \kappa$, for $\theta_L \to 0$ we have

$$\Pi^D = A(\beta\theta_H)((1-\beta)\theta_H - \kappa) < A(\beta\theta_H)(\theta_H - \beta\theta_H) = A(\beta\theta_H)((1-\beta)\theta_H) \leq \Pi^{C_H}.$$

A lower bound for the firm's optimal profit under D is given for $w_H = \beta\Delta\theta + \kappa$, i.e.,

$$\theta_L + A(\beta\Delta\theta)((1-\beta)\Delta\theta - \kappa),$$

which approaches θ_H as $\theta_L \to \theta_H$. Hence, the optimal profit under D exceeds Π^{C_H} for sufficiently large values of θ_L. Thus, as Π^D is monotonically increasing in θ_L irrespective of whether the no-collusion constraint binds or not, there exists a threshold $\hat{\theta}_L \in (0, \theta_H)$ such that $\Pi^D < (>)\, \Pi^{C_H}$ iff $\theta_L < (>)\, \hat{\theta}_L$. In Proposition 7, $\bar{\theta}_L = \min\{\tilde{\theta}_L, \hat{\theta}_L\}$. ■

References

Ackerberg, D. and M. Botticini (2002). Endogenous matching and the empirical determinants of contract form. *Journal of Political Economy 110*, 564–591.

Anderson, S. and R. Renault (2006). Advertising content. *American Economic Review 96*, 93–113.

Bagwell, K. (2007). The economic analysis of advertising. In M. Armstrong and R. Porter (Eds.), *Handbook of Industrial Organization*, Volume 3, pp. 1701–1844. Elsevier: Amsterdam.

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.

Bhardwaj, P. (2001). Delegating pricing decisions. *Marketing Science 20*, 143–169.

Brickley, J. A., C. W. Smith, and J. L. Zimmerman (2009). *Managerial Economics and Organizational Architecture*. McGraw-Hill/Irwin New York.

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.

Desai, P. S. and D. Purohit (2004). "Let me talk to my manager": Haggling in a competitive environment. *Marketing Science 23*(2), 219–233.

DeVaro, J. and F. A. Kurtulus (2010). An empirical analysis of risk, incentives and the delegation of worker authority. *Industrial and Labor Relations Review 63*(4), 637–657.

DeVaro, J. and S. Prasad (2015). The relationship between delegation and incentives across occupations: Evidence and theory. *Journal of Industrial Economics 63*(2), 279–312.

Frank, D. H. and T. Obloj (2014). Firm-specific human capital, organizational incentives, and agency costs: Evidence from retail banking. *Strategic Management Journal 35*, 1279–1301.

Frenzen, H., A.-K. Hansen, M. Krafft, M. K. Mantrala, and S. Schmidt (2010). Delegation of pricing authority to the sales force: An agency-theoretic perspective of its determinants and impact on performance. *International Journal of Research in Marketing 27*, 58–68.

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 (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization 7, Special Issue*, 24–52.

Hong, B., L. Kueng, and M.-J. Yang (2015). Estimating management practice complementarity between decentralization and performance pay. Technical report, National Bureau of Economic Research.

Jia, Y. and P. van Veen-Dirks (2015). Control system design at the production level. *Working Paper*.

Jindal, P. and P. Newberry (2015). Price negotiation and bargaining costs. Working Paper.

Joseph, K. (2001). On the optimality of delegating pricing authority to the sales force. *Journal of Marketing 65*, 62–70.

Kräkel, M. and A. Schöttner (2016). Optimal sales force compensation. *Journal of Economic Behavior and Organization 126*, 179–195.

Laffont, J.-J. and D. Martimort (2002). *The Theory of Incentives: The Principal-Agent Model*. Princeton University Press: Princeton and Oxford.

Lal, R. (1986). Technical note – delegating pricing responsibility to the salesforce. *Marketing Science 5*, 159–168.

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.

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

Lim, N. and S. H. Ham (2014). Relationship organization and price delegation: An experimental study. *Management Science 60*, 586–605.

Lo, D. H.-F., W. Dessein, M. Ghosh, and F. Lafontaine (2016). Price delegation and performance pay: Evidence from industrial sales forces. *Journal of Law, Economics, and Organization 32*, 508–544.

Mishra, B. K. and A. Prasad (2004). Centralized pricing versus delegating pricing to the salesforce under information asymmetry. *Marketing Science 23*, 21–27.

Mishra, B. K. and A. Prasad (2005). Delegating pricing decisions in competitive markets with symmetric and asymmetric information. *Marketing Science 24*, 490–497.

Nagar, V. (2002). Delegation and incentive compensation. *The Accounting Review 77*(2), 379–395.

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy 78*, 311–329.

Nelson, P. (1974). Advertising as information. *Journal of Political Economy 82*, 729–754.

Ohlendorf, S. and P. Schmitz (2012). Repeated moral hazard and contracts with memory: The case of risk-neutrality. *International Economic Review 53*, 433–452.

Owan, H., T. Tsuru, and K. Uehara (2015). Incentives and gaming in a nonlinear compensation scheme: Evidence from north american auto dealership transaction data. *Evidence-based HRM: a Global Forum for Empirical Scholarship 3*(3), 222–243.

Prendergast, C. (2002). The tenuous trade-off between risk and incentives. *Journal of Political Economy 110*(5), 1071–1102.

Riley, J. and R. Zeckhauser (1983). Optimal selling strategies: When to haggle, when to hold firm. *Quarterly Journal of Economics 98*(2), 267–289.

Schöttner, A. (2016). Optimal sales force compensation in dynamic settings: Commissions vs. bonuses. *Management Science Articles in Advance.*

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? *Marketing Science 33*, 796–808.

Stephenson, P. R., W. L. Cron, and G. L. Frazier (1979). Delegating pricing authority to the sales force: The effects on sales and profit performance. *Journal of Marketing 43*, 21–28.

Tirole, J. (1986). Hierarchies and bureaucracies: On the role of collusion in organizations. *Journal of Law, Economics, & Organization 2*, 181–214.

Tirole, J. (1992). Collusion and the theory of organizations. In J.-J. Laffont (Ed.), *Advances in Economic Theory, Sixth World Congress*, Volume 2, pp. 152–213. Cambridge University Press: Cambridge.

Wang, R. (1995). Bargaining versus posted-price selling. *European Economic Review 39*, 1747–1764.

Weinberg, C. B. (1975). An optimal commission plan for salesmen's control over price. *Management Science 21*, 937–943.

Weitz, B. A., S. B. Castleberry, and J. F. Tanner (1998). *Selling: Building Relationships*. Boston: Irwin/McGraw-Hill.

Wulf, J. (2007). Authority, risk, and performance incentives: Evidence from division manager positions inside firms. *Journal of Industrial Economics 55*(1), 169–196.

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.

Supplementary Material

We consider an arbitrary delegation contract $(w_{\varnothing}, w(p), P)$ where P contains at least two elements or equivalently $p_L < p_H$. That is, under a delegation contract, the agent has some pricing authority. We show that each such delegation contract is weakly dominated by a contract where the only feasible prices are $p_L = \theta_L$ and $p_H = \theta_H$. Thus, when analyzing delegation contracts, we can w.l.o.g. restrict attention to $P = \{\theta_L, \theta_H\}$.

Preliminary remarks:

- In analogy to the centralization case, the agent's limited liability constraint implies the participation constraint, which can be skipped in the following. Moreover, $w_{\varnothing} = 0$ is optimal as any positive wage would only increase the firm's labor costs and decrease the agent's search incentives.

- We can restrict attention to $p_H \leq \theta_H$ because a price offer larger than θ_H would prevent any sale.

- Let ignore wages $w(p)$ for a moment. If there is a price $p \in P$ with $p > p_L$ and $p \leq \theta$, the agent prefers offering p and accepting the kickback to offering the price p_L.

  - If the agent finds a θ_H-customer, he prefers offering p_H to offering p_L.
  - If the agent finds a θ_L-customer, the highest price that the agent can offer is

  $$\hat{p}_L := \max\{p : p \in P \text{ and } p \leq \theta_L\}.$$

  If $\hat{p}_L > p_L$, the agent prefers offering $\hat{p}_L$ to offering p_L to a θ_L-customer.

- Let $\bar{p}_i$ denote the price that maximizes $w(p)$ in case a θ_i-customer $(i = L, H)$ is found,[35]

  $$\bar{p}_i := \arg\max_{p \in P,\, p \leq \theta_i} w(p) \quad \text{for } i = L, H.$$

  - If the agent finds a θ_H-customer, he either offers $\bar{p}_H$ and rejects the kickback or he offers p_H and accepts the kickback. The agent will offer $\bar{p}_H$ iff

  $$w(\bar{p}_H) - \kappa \geq w(p_L) + \beta \cdot (p_H - p_L). \qquad \text{(nc-H)}$$

[35] If this price is not unique, let $\bar{p}_i$ denote the highest price that maximizes $w(p)$.

– If the agent finds a θ_L-customer, two cases need to be distinguished: (i) $p_L = \hat{p}_L$ and (ii) $p_L < \hat{p}_L$. In case (i), there is only one price the agent can offer to the customer and the customer cannot offer a positive kickback. In case (ii), the agent either offers $\bar{p}_L$ and rejects the kickback or he offers $\hat{p}_L$ and accepts the kickback. He will offer $\bar{p}_L$ iff

$$w(\bar{p}_L) - \kappa \geq w(p_L) + \beta \cdot (\hat{p}_L - p_L). \tag{nc-L}$$

Hence, under any given contract $(w_\varnothing, w(p), P)$, only the prices p_L, $\bar{p}_L$, $\hat{p}_L$, $\bar{p}_H$, p_H and the corresponding wages are relevant for the game. It is thus w.l.o.g. to eliminate all other prices from P. Note that

$$p_L \leq \bar{p}_L \leq \hat{p}_L \leq \theta_L \quad \text{and} \quad p_L \leq \bar{p}_H \leq p_H \leq \theta_H.$$

Let the corresponding wages $w(p)$ be denoted by

$$w_L := w(p_L),\ \bar{w}_L := w(\bar{p}_L),\ \hat{w}_L := w(\hat{p}_L),\ \bar{w}_H := w(\bar{p}_H),\ w_H := w(p_H).$$

We need to show that we can restrict attention to contracts with $p_L = \bar{p}_L = \hat{p}_L = \theta_L$ and $\bar{p}_H = p_H = \theta_H$.

A delegation contract $(w_\varnothing, w(p), P)$ induces one of the following four cases at the last stage of the game, which we distinguish in the following:

(I) Agent rejects kickback from both types $\theta \in \{\theta_L, \theta_H\}$, i.e., (nc-H) and (nc-L) hold.

(II) Agent accepts kickback from both types $\theta \in \{\theta_L, \theta_H\}$, i.e., (nc-H) and (nc-L) do not hold.

(III) Agent accepts kickback from type θ_H but not from type θ_L, i.e., (nc-H) does not hold but (nc-L) holds.

(IV) Agent accepts kickback from type θ_L but not from type θ_H, i.e., (nc-H) holds but (nc-L) does not.

First, we show that if $\hat{p}_L > \bar{p}_L$ ($p_H > \bar{p}_H$) under a given contract, then we can eliminate either $\hat{p}_L$ or $\bar{p}_L$ (either p_H or $\bar{p}_H$) from P without affecting the allocation. Consequently, the new "reduced" payoff-equivalent contract has (at most) three prices, p_L, $\hat{p}_L$, and p_H, as the wage-maximizing prices $\bar{p}_L$ and $\bar{p}_H$ coincide with one of the prices p_L, $\hat{p}_L$, or p_H.

- Suppose $(w_\varnothing, w(p), P)$ induces **Case (I)**: The firm receives the prices $\bar{p}_L$ or $\bar{p}_H$ and pays the corresponding wages $\bar{w}_L$ or $\bar{w}_H$. If $\hat{p}_L > \bar{p}_L$ ($p_H > \bar{p}_H$), eliminating $\hat{p}_L$ (p_H) from P does not affect the allocation. It just relaxes the no-collusion constraint (nc-L) (or (nc-H)).

- Suppose $(w_\varnothing, w(p), P)$ induces **Case (II)**: The firm always receives the price p_L and pays the wage w_L. If $\hat{p}_L > \bar{p}_L$ ($p_H > \bar{p}_H$), eliminating $\bar{p}_L$ ($\bar{p}_H$) from P and setting $\hat{w}_L = 0$ ($w_H = 0$), does not affect the allocation.

- Suppose $(w_\varnothing, w(p), P)$ induces **Case (III)**: If $\hat{p}_L > \bar{p}_L$ ($p_H > \bar{p}_H$), eliminating $\hat{p}_L$ (eliminating $\bar{p}_H$ and setting $w_H = 0$), does not affect the allocation.

- Suppose $(w_\varnothing, w(p), P)$ induces **Case (IV)**: If $\hat{p}_L > \bar{p}_L$ ($p_H > \bar{p}_H$), eliminating $\bar{p}_L$ and setting $\hat{w}_L = 0$, (eliminating p_H) does not affect the allocation.

Hence, w.l.o.g. we can restrict attention to $P = \{p_L, \hat{p}_L, p_H\}$. We now show that w.lo.g. we can restrict attention to $p_L = \hat{p}_L = \theta_L$ and $p_H = \theta_H$. Given a contract $(w_\varnothing, w(p), P)$ and a customer has been found, let Y denote the agent's expected net payoff from interacting with the customer and being paid by the firm. Let $A(Y)$ denote the probability that the agent finds a customer given Y.[36]

Case (I). Suppose the firm wants to design a contract that induces Case (I). The no-collusion constraints (nc-H) and (nc-L) need to hold and imply that $w_L < \bar{w}_L \in \{\hat{w}_L, w_H\}$ and $w_L < \bar{w}_H \in \{\hat{w}_L, w_H\}$. Hence, $w_L < \min\{\hat{w}_L, w_H\}$. A delegation contract with $\hat{w}_L > w_H$ is always weakly dominated by a centralization contract: Such a contract would imply $\bar{w}_L = \bar{w}_H = \hat{w}_L$. The agent thus offers $\hat{p}_L$ to both customer types, and the firm is weakly better off by eliminating p_L and p_H from P.

Hence, we now consider the case $\hat{w}_L \leq w_H$, implying that $\bar{w}_L = \hat{w}_L$ and $\bar{w}_H = w_H$. The agent offers p_H to a θ_H-customer and $\hat{p}_L$ to a θ_L-customer. The proof proceeds as follows: First, we show that $p_L \neq \theta_L$ cannot be optimal. Second, we show that $p_H < \theta_H$ cannot be optimal either. It then follows that $p_L = \hat{p}_L = \theta_L$ and $p_H = \theta_H$.

The proof is by contradiction. First, suppose that $p_L < \theta_L$. Because $p_L < p_H$, a θ_H-customer will always offer a kickback. By contrast, a θ_L-customer will offer a kickback iff

[36] Recall from Section 3.1 that $A(\cdot) := g(e(\cdot))$ with e denoting the inverse function of $c'(a)/g'(a)$.

$p_L < \hat{p}_L$. We thus have[37]

$$Y = q(w_H - \kappa) + (1-q)(\hat{w}_L - \mathbb{I}_{\{p_L < \hat{p}_L\}}\kappa).$$

Let $Y_f := qw_H + (1-q)\hat{w}_L$ denote the expected payment that the agent obtains from the firm and $Y_c := Y - Y_f$ the payoff that the agent receives from interacting with the customer. Given P, the firm then chooses $w(p)$ to solve

$$\max A(Y_f + Y_c)(qp_H + (1-q)\hat{p}_L - Y_f)$$

$$\text{subject to } w_H - \kappa \geq w_L + \beta(p_H - p_L), \tag{26}$$

$$\hat{w}_L - \kappa \geq w_L + \beta(\hat{p}_L - p_L). \tag{27}$$

As long as $p_L < \hat{p}_L$, the firm is weakly better off by increasing p_L because doing so relaxes (26) and (27) without being payoff relevant. Hence, $p_L = \hat{p}_L$ and there is no scope for collusion with a θ_L-customer anymore, which increases Y and Y_c. Therefore, (27) can be neglected. Now, if $\hat{p}_L < \theta_L$, the firm will benefit from increasing $\hat{p}_L$ because it then receives a higher price from a θ_L-customer. It follows that $p_L = \hat{p}_L = \theta_L$ dominates $p_L < \theta_L$.

Now suppose $\theta_L < p_L$. Hence, if a θ_L-customer is found, the game will end. We have $Y_f = qw_H$, $Y_c = -q\kappa$ and, given P, the firm chooses $w(p)$ to solve

$$\max A(Y_f + Y_c)(qp_H - Y_f) \quad \text{subject to (26).}$$

As long as $p_L < p_H$, the firm is thus weakly better off by increasing p_L to relax (26). Hence, centralization with $P = \{p_H\}$ dominates delegation with $\theta_L < p_L < p_H$. It follows that, under a delegation contract, we must have $p_L = \theta_L$.

Finally, we show by contradiction that $p_H < \theta_H$ cannot be optimal.[38] Suppose that $p_H < \theta_H$. Because $p_L = \theta_L$, we have $Y_f = qw_H + (1-q)w_L$ and $Y_c = -q\kappa$. Given P, the firm chooses $w(p)$ to solve

$$\max A(Y_f + Y_c)(qp_H + (1-q)p_L - Y_f) \quad \text{subject to (26)}$$

[37] The indicator function $\mathbb{I}_{\{p_L < \hat{p}_L\}}$ takes the value 1 if $p_L < \hat{p}_L$, and the value 0 otherwise. It is used to jointly analyze two subcases.

[38] Intuition: Lowering p_H below θ_H decreases the kickback that the θ_H-customer can offer. However, the firm also obtains a lower price from this customer. The latter effect always dominates the former.

Defining $\Delta w := w_H - w_L$, we obtain

$$\max_{w_L,\Delta w} A(w_L + q\Delta w - q\kappa)(qp_H + (1-q)p_L - (w_L + q\Delta w))$$
$$\text{subject to } \Delta w \geq \kappa + \beta(p_H - p_L).$$

An optimal solution is $w_L = 0$ and

$$\Delta w = w_H = \begin{cases} \kappa + \beta(p_H - \theta_L) & \text{if } \Gamma(\kappa + \beta(p_H - \theta_L)) \leq 0, \\ w_H^* & \text{otherwise.} \end{cases}$$

where

$$\Gamma(w_H) := qA'(q(w_H - \kappa))(qp_H + (1-q)p_L - qw_H) - qA(q(w_H - \kappa))$$

describes the first derivative of the firm's strictly concave objective function with respect to w_H, and the interior solution w_H^* is described by $\Gamma(w_H^*) = 0$. The firm's profit thus is

$$\Pi = \begin{cases} A(q\beta(p_H - \theta_L))\times & \\ \quad (qp_H + (1-q)p_L - q(\kappa + \beta(p_H - \theta_L))) & \text{if } \Gamma(\kappa + \beta(p_H - \theta_L)) \leq 0, \\ A(q(w_H^* - \kappa))(qp_H + (1-q)p_L - qw_H^*) & \text{otherwise.} \end{cases}$$

The profit is always increasing in p_H.[39]

Case (II). Suppose the firm wants to design a contract that induces Case (II), i.e., the constraints (nc-H) and (nc-L) do not hold. If the agent sells the product, the firm always receives p_L and pays w_L to the agent. W.l.o.g., the firm can stipulate $w_H = \hat{w}_L = 0$ to ensure that (nc-H) and (nc-L) do not hold. Because $p_H \leq \theta_H$, the agent always sells to a θ_H-customer. Iff $p_L \leq \theta_L$, the agent also sells to a θ_L-customer. We have

$$Y = q[w_L + \beta(p_H - p_L)] + (1-q)\mathbb{I}_{\{p_L \leq \theta_L\}}[w_L + \beta(\hat{p}_L - p_L)].$$

Given P, the firm chooses w_L to maximize

$$A(Y)(q(p_L - w_L) + (1-q)\mathbb{I}_{\{p_L \leq \theta_L\}}(p_L - w_L)).$$

[39] In the second line, w_H^* is a function of p_H. However, as for the interior solution w_H^* the first-order condition $\Gamma(w_H^*) = 0$ holds, the Envelope Theorem applies and we only have to care about the direct effect of p_H on Π.

To maximize the incentive effect via the kickback paid by a θ_H-customer, the firm sets p_H as high as possible, i.e., $p_H = \theta_H$.

We now suppose that $p_L \neq \theta_L$ and show that the firm can then always increase its profit by raising p_L.[40] Note that we can write the firm's problem of choosing the optimal wage w_L as

$$\max_{Y_f} A(Y_f + Y_c)(R - Y_f),$$

where Y_f denotes the expected payment from the firm to the agent and $Y_c := Y - Y_f$, i.e., Y_c is the agent's payoff stemming from the interaction with the customer. Moreover, R denotes the firm's expected revenue. Specifically, we have

$$\begin{aligned} Y_f &= q w_L + (1-q)\mathbb{I}_{\{p_L < \theta_L\}} w_L \\ Y_c &= q\beta(p_H - p_L) + (1-q)\mathbb{I}_{\{p_L < \theta_L\}} \beta(\hat{p}_L - p_L) \\ R &= q p_L + (1-q)\mathbb{I}_{\{p_L < \theta_L\}} p_L \end{aligned}$$

Let $\Gamma(Y_f) := A'(Y_f + Y_c)(R - Y_f) - A(Y_f + Y_c)$ denote the first derivative of the firm's strictly concave objective function with respect to Y_f. If $\Gamma(0) \leq 0$, then the firm's optimal expected payment is $Y_f^* = 0$. Otherwise, $Y_f^* > 0$ and described by $\Gamma(Y_f^*) = 0$. The firm's profit is

$$\Pi(p_L) = \begin{cases} A(Y_c)R & \text{if } \Gamma(0) \leq 0, \\ A(Y_f^* + Y_c)(R - Y_f^*) & \text{otherwise.} \end{cases}$$

For the case $\Gamma(0) \leq 0$, we obtain[41]

$$\frac{\partial \Pi}{\partial p_L} = A'(Y_c)\underbrace{\frac{\partial Y_c}{\partial p_L}}_{-} R + A(Y_c)\underbrace{\frac{\partial R}{\partial p_L}}_{+} > 0 \quad \Leftrightarrow \quad \frac{A(Y_c)}{RA'(Y_c)} > \frac{-\frac{\partial Y_c}{\partial p_L}}{\frac{\partial R}{\partial p_L}}.$$

Note that $\Gamma(0) \leq 0$ is equivalent to $\frac{A(Y_c)}{RA'(Y_c)} \geq 1$. In addition, because $q, \beta < 1$, we obtain $-\frac{\partial Y_c}{\partial p_L} < \frac{\partial R}{\partial p_L}$ for all possible cases $p_L > \theta_L$, $p_L < \hat{p}_L \leq \theta_L$, and $p_L = \hat{p}_L < \theta_L$. Hence, $\frac{\partial \Pi}{\partial p_L} > 0$.

[40] Intuition: If $p_L > \theta_L$, the firm benefits from increasing p_L up to p_H. This decreases the kickback and hence the agent's incentives associated with the kickback, but this effect is dominated by the higher price the firm can collect. If $p_L < \theta_L$, the firm benefits from increasing p_L up to θ_L. Again, the higher price the firm can collect dominates the agent's lower incentives via smaller kickbacks.

[41] Recall that the indicator function only distinguishes the two subcases $p_L > \theta_L$ and $p_L < \theta_L$ so that there is no discontinuity when computing the derivative.

For the case $\Gamma(0) > 0$, we obtain[42]

$$\frac{\partial \Pi}{\partial p_L} = A'(Y_f^* + Y_c)\frac{\partial Y_c}{\partial p_L}(R - Y_f^*) + A(Y_f^* + Y_c)\frac{\partial R}{\partial p_L} > 0$$
$$\overset{R-Y_f^*=\frac{A(\cdot)}{A'(\cdot)}}{\iff} A(Y_f^* + Y_c)\left(\frac{\partial Y_c}{\partial p_L} + \frac{\partial R}{\partial p_L}\right) > 0,$$

which is again true because $-\frac{\partial Y_c}{\partial p_L} < \frac{\partial R}{\partial p_L}$. Hence, $p_L < \theta_L$ is dominated by $p_L = \theta_L$ and $\theta_L < p_L$ is dominated by $p_L = p_H$ (centralization). Consequently, as long as $p_L < p_H$, we must have $p_L = \theta_L$ and $p_H = \theta_H$.

Case (III). Suppose the firm wants to design a contract such that the agent accepts a kickback from a θ_H-customer but rejects a kickback from a θ_L-customer, i.e., (nc-H) does not hold while (nc-L) holds. This case is relevant only if $p_L < \hat{p}_L$.[43] The firm should set $p_H = \theta_H$ to maximize the agent's kickback from a θ_H-customer and hence his effort incentives. To ensure that (nc-H) does not hold, the firm can w.l.o.g. set $w_H = 0$. If the agent encounters a θ_L-customer, he offers $\hat{p}_L$ and sells at this price ($\bar{w}_L = \hat{w}_L > w_L$ must hold to ensure (nc-L)). Hence,

$$Y = q(w_L + \beta(p_H - p_L)) + (1 - q)(\hat{w}_L - \kappa).$$

Given P, the firm chooses $w(p)$ to solve

$$\max A(Y)[qp_L + (1 - q)\hat{p}_L - qw_L - (1 - q)\hat{w}_L]$$
$$\text{subject to } w_L + \beta(p_H - p_L) > w_H - \kappa \tag{28}$$
$$\hat{w}_L - \kappa \geq w_L + \beta(\hat{p}_L - p_L). \tag{29}$$

Because $w_H = 0$, constraint (28) holds. Dropping constraint (29), we obtain a relaxed problem for which we can show that the optimal profit is increasing in p_L as long as $p_L < \hat{p}_L$.[44] Hence, $p_L = \hat{p}_L$ is also optimal for the original problem because (29) is not relevant anymore since a θ_L-customer cannot offer a kickback. Therefore, Case (III) with

[42] Again, we make use of the Envelope Theorem so that the indirect effect of p_L on Π via Y_f^* can be skipped. In addition, $R - Y_f = A(\cdot)/A'(\cdot)$ follows from $\Gamma(Y_f^*) = 0$.

[43] For $p_L = \hat{p}_L$, the analysis of Case (II) shows that $p_L = \theta_L$ and $p_H = \theta_H$ under an optimal contract.

[44] The formal proof is given below (it covers Case (III) and Case (IV)). Intuition: If $p_L = \hat{p}_L$ and a θ_L-customer is found, the firm does not need to prevent collusion. In case a θ_H-customer is found, the kickback and thus the agent's incentives stemming from the kickback are lower, but this effect is dominated by the higher price that the firm obtains for the product (compare Case (II)).

$p_L < \hat{p}_L$ cannot constitute an optimal contract.

Case (IV). Suppose the firm wants to design a contract such that the agent accepts a kickback from a θ_L-customer but rejects a kickback from a θ_H-customer, i.e., (nc-H) holds while (nc-L) does not hold. This case is relevant only if $p_L < \hat{p}_L$.[45] If the agent encounters a θ_L-customer, he offers $\hat{p}_L$, accepts the kickback and sells for p_L. If the agent encounters a θ_H-customer, he offers $\bar{p}_H \in \{\hat{p}_L, p_H\}$ and receives $\bar{w}_H \in \{\hat{w}_L, w_H\}$. Hence,

$$Y = q(\bar{w}_H - \kappa) + (1 - q)(w_L + \beta(\hat{p}_L - p_L)).$$

Given P, the firm chooses $w(p)$ to solve

$$\begin{aligned} &\max A(Y)(q\bar{p}_H + (1 - q)p_L - q\bar{w}_H - (1 - q)w_L) \\ &\text{subject to } \bar{w}_H - \kappa \geq w_L + \beta(p_H - p_L) && (30) \\ &w_L + \beta(\hat{p}_L - p_L) > \hat{w}_L - \kappa. && (31) \end{aligned}$$

If $\bar{w}_H = \hat{w}_L$, then (30) and (31) are in contradiction. Hence, $\bar{w}_H = w_H$ and $\bar{p}_H = p_H$. W.l.o.g., the firm can set $\hat{w}_L = 0$ to ensure that constraint (31) holds. To maximize the incentive effect via the kickback obtained from a θ_L-customer, the firm should not exclude θ_L from P, i.e., $\hat{p}_L = \theta_L$. Neglecting (30), we can show that the firm's optimal profit is increasing in p_L as long as $p_L < \theta_L$. Because increasing p_L also relaxes constraint (30), $p_L < \theta_L$ is dominated by $p_L = \theta_L$. Hence, Case (IV) cannot be optimal.

Formal proof that the firm's profits are strictly increasing in p_L as long as $p_L < \hat{p}_L$ in Case (III) and (IV) (neglecting constraints (29) and (30), respectively). As in Case (II), we can write the firm's objective function as

$$\max_{Y_f} A(Y_f + Y_c)(R - Y_f),$$

where in Case (III) we have

$$\begin{aligned} Y_f &= qw_L + (1 - q)\hat{w}_L, \\ Y_c &= q\beta(p_H - p_L) - (1 - q)\kappa, \\ R &= qp_L + (1 - q)\hat{p}_L, \end{aligned}$$

[45] For $p_L = \hat{p}_L$, the analysis of Case (I) shows that $p_L = \theta_L$ and $p_H = \theta_H$ under an optimal contract.

and in Case (IV) we have

$$
\begin{aligned}
Y_f &= qw_H + (1-q)w_L, \\
Y_c &= -q\kappa + (1-q)\beta(\hat{p}_L - p_L), \\
R &= qp_H + (1-q)p_L.
\end{aligned}
$$

We can now proceed exactly as in Case (II): Define $\Gamma(Y_f) := A'(Y_f+Y_c)(R-Y_f)-A(Y_f+Y_c)$. If $\Gamma(0) \leq 0$, then the firm's optimal expected payment will be $Y_f^* = 0$. Otherwise, $Y_f^* > 0$ and described by $\Gamma(Y_f^*) = 0$. The firm's profit is

$$
\Pi = \begin{cases} A(Y_c)R & \text{if } \Gamma(0) \leq 0, \\ A(Y_f^* + Y_c)(R - Y_f^*) & \text{otherwise.} \end{cases}
$$

For the case $\Gamma(0) \leq 0$, we obtain

$$
\frac{\partial \Pi}{\partial p_L} = A'(Y_c)\underbrace{\frac{\partial Y_c}{\partial p_L}}_{-} R + A(Y_c)\underbrace{\frac{\partial R}{\partial p_L}}_{+} > 0 \quad \Leftrightarrow \quad \frac{A(Y_c)}{RA'(Y_c)} > \frac{-\frac{\partial Y_c}{\partial p_L}}{\frac{\partial R}{\partial p_L}}.
$$

Note that $\Gamma(0) \leq 0$ is equivalent to $\frac{A(Y_c)}{RA'(Y_c)} \geq 1$. Moreover, $-\frac{\partial Y_c}{\partial p_L} < \frac{\partial R}{\partial p_L}$ in both Case (III) and Case (IV). Hence, $\frac{\partial \Pi}{\partial p_L} > 0$.

For the case $\Gamma(0) > 0$, again the Envelope Theorem applies and we obtain

$$
\begin{aligned}
\frac{\partial \Pi}{\partial p_L} &= A'(Y_f^* + Y_c)\frac{\partial Y_c}{\partial p_L}(R - Y_f^*) + A(Y_f^* + Y_c)\frac{\partial R}{\partial p_L} > 0 \\
&\overset{R-Y_f^* = \frac{A(\cdot)}{A'(\cdot)}}{\Longleftrightarrow} A(Y_f^* + Y_c)\left(\frac{\partial Y_c}{\partial p_L} + \frac{\partial R}{\partial p_L}\right) > 0,
\end{aligned}
$$

which is again true because $-\frac{\partial Y_c}{\partial p_L} < \frac{\partial R}{\partial p_L}$.