

(Mis-)selling through Agents

Roman Inderst* Marco Ottaviani†

April 2007

Preliminary Draft

Abstract

We study deceptive selling (or misselling) through a firm's agents. Agents who are paid to originate sales are biased towards making deceptive statements or providing wrong advice so as to increase sales volume. We ask first what affects firms' costs to ensure compliance with a given standard, thereby determining when firms become more permissive towards misselling. We then show how by taking into consideration the underlying agency problem the optimal policy response must depend on the degree of competition, the organization of the sales process, the nature of compliance costs, and the transparency of the commission structure.

*University of Frankfurt and London School of Economics, r.inderst@lse.ac.uk.

†London Business School, mottaviani@london.edu.

1 Introduction

When purchasing complex or unfamiliar products, consumers often rely on sales advice. This has long been recognized in the vast literature on credence goods. Motivated often by the picture of a professional service provider such as a doctor, this literature has by and large abstracted from the fact that frequently the person who is supposed to provide advice acts as a self-interested agent of the seller. In the simplest case, the agent could be a seller's employee, though in many industries firms increasingly outsource their sales activity, including thus the function to advise potential customers. Casual evidence suggests that such agency relationships can increase the risk of deceptive selling or misselling. For instance, in relation to the recent turmoil in the US subprime mortgage market the Economist observes that “[m]any [customers] appear to have been encouraged to take out loans by brokers more bothered about their fees than their clients' ability to repay their debts” and that “[m]any of the riskiest mortgages were made by independent, non-bank lenders.”¹

Our model analyzes the possibility of misselling through the lenses of this agency relationship between the firm and its salesforce or the independent advisors it relies on. We argue that the risk of misselling is particularly acute when the firm hires the same agents both to “market” the good or service, e.g., by prospecting new customers, and to provide advice. In particular, if it is necessary or simply profitable for the firm to provide high-powered incentives for the first task, for instance as competition makes it hard to locate potential customers, then this will increasingly bias the agent towards either inflating the value of the good through deceptive statements or recommending purchase also to customers for whom the good is unsuitable.

Under the *respondeat superior* doctrine, a firm selling through its own employees is vicariously liable for negligent acts of these subordinates within the course of employment. However, even if misselling takes place through independent intermediaries, the firm may still risk being sued or, in case of a regulated industry, losing its licence and facing additional penalties. For instance, though independent financial advisers and mortgage brokers play an essential part in the UK's retail finance sector, in its recent misselling scandals the bulk of legal costs and compensation payments was shouldered by the “big players” whose products were sold.²

¹‘The trouble with the housing market’ (page 11) and ‘America's housing market’ (page 79), March 24, 2007.

²The two biggest scandals involved the alleged misselling of private pensions, which after a full review in 1994 reportedly led to a total compensation in excess of £12 billion, and the alleged misselling of endowment mortgages,

Though firms should thus have a vital interest in ensuring that their agents show full compliance with the required standards, this may involve costly internal reviews and may, as shown in our model, leave agents with high rents. These costs are lower if it is possible to use contingent (sales) commissions tying, for instance, the overall compensation to whether dissatisfied customers cancelled a contract early or not. On the other hand, if the nature of the product much restricts the firm's contractual instruments, then only compliance to sufficiently low standards may be feasible for the firm. Also, the organization of the sales process, which may naturally differ depending on products and firms, determines the level of misselling that the firm will wish to tolerate. If agents are relieved from the task of "marketing" the goods, e.g., as they are only advising incoming clients in a bank's local branch, then misselling will represent less of a problem.³ On the other hand, if the same agent must both contact new customers and advise them, then the firm will face a trade-off in providing incentives along both dimensions.

The main application of our basic model is to study the role and scope of policies to curb misselling. We show how a consideration of the underlying agency problem is important both when defining the targeted standard, as expressed in the expected level of misselling that would then still prevail, and when determining the optimal policy response to achieve this target. Even only slightly higher standards may, through the underlying agency problem, involve much higher compliance costs or may even not be feasible for firms to comply with. In addition, they may backfire by curbing the incentives that firms provide for their salesforce, thereby leading to less or slower penetration of the market, which may cause consumer detriment, in particular, for new and innovative products that lack awareness. Furthermore, we show that the respective policymaker or regulator must monitor firms more or less closely and must impose higher or lower penalties depending, amongst other things, on the intensity of competition or on how the sales process is organized in the industry. In addition, besides imposing stiffer penalties, a higher standard can also be achieved by forcing firms to make transparent to customers the commissions that they pay to their agents.

Though the need for sales advice is particularly acute in the case of financial services, such as pensions and mortgages (Campbell, 2006), our model and its insights apply more widely. In

where compensations are still being made and more than £2 billion has already been paid out to policyholders in redress.

³In fact, we find that in the former case a firm may even, from a social perspective, "under-sell" through discouraging its agents too much from selling to a wider range of customer "types".

particular, as we make clear below, our model equally applies to situations in which agents are tempted to inflate the perceived value of the good or service by making deceptive statements. For instance, a salesperson may praise some features but hide others when trying to convince a client to switch to a particular call plan or utility contract, both of which may be sufficiently complex products so as to make such deception successful.⁴

To the best of our knowledge, our focus on the agency problem is novel to the literature on credence goods (see Darby and Karni, 1973, and the recent review by Dulleck and Kerschbamer, 2006).⁵ Our assumption that the principal (the firm) bears responsibility for the misdeeds of his agents has been investigated more broadly in the literature on vicarious liability (see, e.g., Pitchford, 1995, and Che and Spier, 2006). Furthermore, our insight that the firm faces a trade-off when trying to incentivize the agent along the two tasks of selling and providing advice follows the multi-task literature (Holmström and Milgrom, 1991). More closely related, the variant of our model in which the agent is expected to provide advice and has thus an information advantage vis-a-vis both the principal (the firm) and the customer is related to the problem studied in Levitt and Snyder (1997). There, after exerting effort an agent obtains (interim) private information, which the principal may use to discontinue the agent's project.⁶

The rest of this paper is organized as follows. Section 2 introduces the fundamental agency problem in the most basic model. This is extended and put to various applications in Section 3. Section 4 provides some additional discussion of the basic model. Section 5 concludes.

2 Introducing the Agency Problem

2.1 Model

Consider a firm that sells a single product through an agent, be it either an employee or another company specializing in distribution. By exerting sales effort at private disutility $c_S > 0$, the

⁴See, for instance, the discussion by the UK telecommunication regulator in 'Migration, Switching and Mis-selling', Consultation Document, Ofcom, 16 February 2006, <http://www.ofcom.org.uk/consult/condocs/migrations/migrations.pdf>.

⁵See Bolton et al.'s (2005) analysis of the effect of competition on the incentives for information provision in the market for retail financial services.

⁶For other models of delegated expertise see Lambert (1986), Demski and Sappington (1987), and Lewis and Sappington (1997).

agent obtains a contact with a potential customer with probability $\mu > 0$. We can think of the agent as either contacting previous clients or prospecting new clients.⁷

A key ingredient of the model is that the agent is also expected to give advice to the customer. To capture the role of advice formally, we stipulate for the moment that there are two types of customers, $\theta = l, h$, who could derive the respective gross utilities $u_h > u_l$ from consuming the good. The prior probability that $\theta = h$ is given by $0 < \pi < 1$, which is also the only information that the customer has. In contrast, the agent either already knows or can find out how suitable the good is for the respective customer. For simplicity, though this does not affect results qualitatively, we stipulate that when incurring private costs of “advice” $c_A \geq 0$ the agent perfectly learns the customer’s type. As this is not key to our basic analysis, we set $c_A = 0$ until Section 3.

Contracts that condition directly on the agent’s effort are assumed not to be feasible due to problems of observability or verifiability. What is verifiable is, first, the fact whether a sale has been made and, second, following a sale also a signal about whether the product was indeed suitable for the customer. For the moment, we assume that this signal is costless and that it reveals with probability $0 < \psi < 1$ whether a sale was made to a type- l customer.⁸ For instance, a customer who ultimately finds out that the product was not suitable as he realizes only u_l would complain with the firm by contesting the sale with probability q . We have more to say on the interpretation of the signal in the following Section. (Also, Section 4 extends the analysis of the basic model to the case where generating the signal is costly for the firm.)

We also stipulate that the agent’s market wage is zero and that he is protected by limited liability (or, likewise, has zero wealth). While somewhat extreme, our results will continue to hold whenever the limited liability constraint binds. It is then intuitive, and easy to show, that the optimal compensation scheme will specify the lowest possible wage, namely that of zero, in case it was revealed that the agent sold to a type- l customer. This leaves us with two compensation levels for the other two verifiable states, namely a wage of w_1 if no sale was made and a wage of w_2 if following a sale no negative information was received.

⁷We could also lump into c_S the costs of making a “sales pitch” to the customer.

⁸Note that we could likewise stipulate more symmetrically that following a sale to any type of customer an informative but noisy signal $s \in \{l, h\}$ was revealed. Though in this case the expressions are slightly more complicated, all results hold.

2.2 Analysis

In our basic model, where the agent can perfectly discriminate between the two possible types, we stipulate that the firm wants the agent only to sell to high-type customers. As discussed in more detail in Section 4, this could be the case as the firm makes a loss from a contract with a low-type agent. For instance, if the firm sells insurance contracts or loans, then it may only break even with certain types of customers. However, our main focus in this paper is on the case where the firm wants to prevent misselling to low-type customers, for whom the good is not suitable. This could in turn be the case as the firm fears that, otherwise, it will lose reputation or face legal consequences. Again, Section 4 provides here more details.

Suppose that the agent has contacted a customer. We ask now when the agent will indeed advise low-type customers not to buy.⁹ Recall first that in this case the agent realizes the payoff w_1 . Instead, if the agent deviates then this will result in a payoff of zero with probability ψ , while with the residual probability $1 - \psi$ the agent will be able to pocket the payoff w_2 . Consequently, the agent will only act in the firm's interests and give the right advice in case

$$w_1 \geq (1 - \psi)w_2. \quad (1)$$

In words, to ensure that the agent does not sell indiscriminately to all customers, the wage w_1 for not selling must compensate the agent for the foregone wage obtained by a sale to the low-type customer, which would be uncontested with probability $1 - \psi$. Intuitively, condition (20) is easier to satisfy if a deviation, i.e., misselling to a low-type customer, is detected with a higher probability ψ .

So far we have assumed that the agent actually finds it worthwhile to contact a customer in the first place, which comes at costs c_S . If the agent subsequently provides the right advice, then his initial effort will lead to a sale with probability $\mu\pi$, given that a potential customer is located with probability μ and that the fraction of high types is π . As the agent can always ensure himself a wage of w_1 , he will thus only exert effort if

$$\mu\pi(w_2 - w_1) \geq c_S. \quad (2)$$

The bonus $w_2 - w_1$ that the agent realizes when making a sale to a high-type customer must be sufficiently large. Intuitively, condition (2) is easier to satisfy if it is overall more likely that,

⁹As will be immediate, under the optimal contract the agent will always strictly prefer to advise high-type customers to buy. We therefore ignore this constraint for the time being.

after incurring the costs c_S , a sale will be made, in which case the agent earns w_2 instead of only w_1 .

To set up the firm's program, note that we take as given that the firm wants to incentivize the agent not to missell to low-type customers. The firm's objective is then simply to choose the compensation scheme (w_1, w_2) so as to minimize the expected wage costs while satisfying both constraints (1) and (2).

If the two constraints bind, which will hold by optimality, then the agent's expected payoff, which is also his rent, is equal to w_1 . This is intuitive as the agent can also realize w_1 without exerting any sales effort. The firm's total expected wage bill is then just $c_S + w_1$, compensating the agent for his costs of effort and transferring the rent w_1 .

Proposition 1. *To ensure that the agent both exerts effort to contact customers and subsequently advise only type-h customers to buy, the firm incurs the total costs of*

$$C = c_S \left(1 + \frac{1 - \psi}{\psi} \frac{1}{\mu\pi} \right). \quad (3)$$

Proof. The firm's objective is to minimize expected wage costs $\mu\pi w_2 + (1 - \mu\pi)w_1$ subject to the constraints (1) and (2). If none was binding, the firm could profitably deviate by marginally reducing either of the two wages w_1 or w_2 . Suppose now that (1) was not binding, in which case we already know that (2) must bind. After substitution, the firm's expected wage costs become then $c_S + w_1$. As long as (1) was indeed not binding, the firm could then profitably deviate by further reducing w_1 . Having thus established that (1) must always bind, we can now substitute this to obtain the expected wage costs $w_2[1 - \psi + \mu\pi]$. Again, unless (2) binds as well the firm could profitably deviate by further reducing this time w_2 . From the two binding constraints (1) and (2) we finally obtain

$$w_1 = c_S \left(\frac{1 - \psi}{\psi} \frac{1}{\mu\pi} \right).$$

From this, we then obtain (3) for the firm's total wage costs $c_S + w$. **Q.E.D.**

The following comparative results follow next immediately from Proposition 1.

Corollary 1. *The firm's costs of preventing misselling to low-type customers are strictly increasing in the costs of sales effort c_S . In addition, costs strictly higher if it is less likely that the effort results in a sale as both μ and π are lower and if it is less likely that misselling will be detected as ψ is lower.*

To shed more light on Proposition 1 and Corollary 1 together, it is worthwhile to note first that given the discrete nature of (sales) effort, the fact that the firm must pay a rent is entirely due to the agent’s second task, namely the provision of advice to customers. A key property for our analysis is that this rent is strictly increasing in the agent’s costs of effort c_S . Furthermore, the rent is strictly lower if it is easier to contact a customer as μ is higher. Overall, the rent is thus higher the more difficult it is to originate a new customer.

Intuitively, in this case the agent must be paid a higher bonus, or wage differential, $w_2 - w_1$, to still elicit his sales effort. More formally, we have from the binding constraint (2) the requirement that

$$w_2 - w_1 = \frac{1}{\pi} \frac{c_S}{\mu}.$$

But the higher $w_2 - w_1$, the more will the agent subsequently be tempted to missell to a low-type customer. Consequently, the firm must in turn make it more attractive not to sell at all, which is done through an increase in w_1 . To see this formally, we can transform the binding constraint (1) into the requirement that

$$w_1 = (w_2 - w_1) \left(\frac{1 - q}{q} \right).$$

The crux for the firm is now that any increase in w_1 represents a pure transfer to the agent, given that the agent could also realize w_1 by not even trying to contact a customer in the first place.

When specifying the feasible contractual set, we assumed that the firm can reduce to agent’s payoff to zero in response to “negative” information following a sale. Implicitly, this also implies that the firm can fully “claw-back” any bonus or commission for the sale that it either promised or already paid the agent. One case in which this should be, at least to some extent, feasible is when a dissatisfied customer may use his (statutory or contractually agreed) right to cancel early a long-term agreement, say an insurance contract or a savings and investment plan. In this case, instead of paying the agent a lump-sum origination fee up-front, his compensation could be tied more closely to the “performance” of the contract over time.

Such contingent commissions are indeed very common in some financial industries. For instance, Hoyt et al. (2006) document their use in the US insurance industry. An interesting, albeit extreme, example is also provided by the UK’s “home credit” market, which comprises typically small-sum loans for which the repayments are collected personally at the customer’s home. In

this market, where some lenders employ thousands of independent agents, compensation is often tied entirely to collections.¹⁰

More generally, though, the scope to which the firm can sufficiently withhold such commissions and bonuses may be quite limited, in particular if the suitability of the good to the customer is unlikely to show already in the near future: Moreover, while the cancellation of a long-term contract represents a verifiable event on which the firm can condition the payment or withholding of a commission, this should be less so with more general feedback from customers or the outcome from internal quality reviews. The more it is then left to the firm's discretion whether to pay the outstanding balance of some promised origination fee, the more this could give rise to an opportunism problem, this time on the side of the firm.

Suppose now that following a negative signal, the firm can claw back or withhold the fraction α of the bonus $b := w_2 - w_1$ that is paid when a new sale is made. It is now also convenient to denote the above-market (baseline) wage by $w = w_1 > 0$. Consequently, the contract is now described by a pair (w, b) . We have the following result.

Proposition 2. *Suppose the firm can only claw back or withhold the fraction α of the bonus that is paid when a sale was made. Then the firm's total costs from ensuring that the agent both exerts sales effort and advises only type- h customers to buy are*

$$C = c_S \left(1 + \frac{1 - \psi\alpha}{\psi} \frac{1}{\mu\pi} \right) \quad (4)$$

which are strictly higher the lower α .

Proof. As the agent now realizes the payoff

$$(w + b)(1 - \psi) + \psi b(1 - \alpha)$$

when making a sale to a low-type customer, given that then only the fraction α of the bonus b can be clawed back, he will avoid misselling only if

$$\psi w \geq b(1 - \psi\alpha). \quad (5)$$

¹⁰This market was described in detail in the UK Competition Commission's recent report 'Home Credit Market Investigation', 30 November 2006, http://www.competition-commission.org.uk/rep_pub/reports/2006/517homecredit.htm.

The second constraint, namely that the agent exerts sales effort, is obtained from (2) simply by substituting $b = w_2 - w_1$, which obtains

$$\mu\pi b \geq c_S. \quad (6)$$

By an argument analogous to that in the proof of Proposition 1, both constraints (5) and (6) must bind, which together yields

$$w = c_S \left(\frac{1 - \psi\alpha}{\psi} \frac{1}{\mu\pi} \right). \quad (7)$$

From this, we then obtain (4) for the firm's total wage costs $c_S + w$. **Q.E.D.**

If the firm can not at all claim back the bonus, then from (4) the firm's costs are

$$C = c_S \left(1 + \frac{1}{\psi} \frac{1}{\mu\pi} \right). \quad (8)$$

For most of the paper from now on, we choose to work with this case with $\alpha = 0$, though all of our results extend to general α .

3 Applications

3.1 Incentives for (Mis-)Selling

To isolate the working of the agency problem, our basic model specified that there are only two types of customers and that the agent can perfectly discriminate between them. In what follows, we enrich this setting to allow for a more gradual variation in the extent of misselling. Precisely, we do so by specifying that the agent is no longer perfectly informed about the customer's type. Hence, while ultimately the product or service is unambiguously suitable for high-type but not for low-type customers, the situation is less black-and-white for the agent at the time the sale is made.

Extending the model

Precisely, we now stipulate that the agent only observes a noisy signal about the customer's respective type and thus the ultimate suitability of the product or service. The signal $s \in [0, 1]$ is realized according to the type-dependent distribution function G_θ . We assume that G_θ has strictly positive and continuous density $g_\theta(s)$ over $s \in (0, 1)$ and that G_h dominates G_l in the sense of the Maximum Likelihood Ratio Property. That is, higher signals are good news. It is

convenient to stipulate that $g_h(1) > 0$, $g_l(0) > 0$, and $g_h(0) = g_l(1) = 0$ such that the signal is fully informative at the boundaries $s = 0, 1$. Depending on the observed signal, this yields a strictly increasing and continuous posterior probability $q(s) = \Pr[\theta = h \mid s]$ with $q(0) = 0$ and $q(1) = 1$. Moreover, from an ex-ante perspective s is distributed according to some distribution $F(s)$ with density $f(s)$.¹¹

As a first step in the analysis, we restate our previous results from the basic model. The main difference to the basic model is that the firm can now control more gradually the type of advice that it wants the agent to give, namely a range of signals $s \in [s^*, 1]$ for which the agent should advise the customer to buy. The first incentive constraint (1) from the basic model, which ensures that the agent provides the right advice, transforms then to the requirement that at s^* the agent is just indifferent between making a sale or not. (Note that we use here continuity of $q(s)$.) Recall now that a sale generates the bonus b for the agent, which is not clawed back. In addition, the agent will then continue to realize the (above-market) baseline wage of w with probability $1 - \psi [1 - q(s^*)]$, which takes into account that with probability $1 - q(s^*)$ the sale was to a type- l customer and that this subsequently generates a negative signal with probability ψ . As the agent can realize w without a sale, we then have the requirement that

$$b = w\psi [1 - q(s^*)]. \quad (9)$$

Note next that the overall probability that the agent makes a sale is now μ times $1 - F(s^*)$. Moreover, whenever he makes a sale for $s \geq s^*$ his expected payoff is now just $b + w [1 - \psi [1 - q(s^*)]]$, taking again into account that the customer's type is not perfectly known at this stage. As the agent can still realize the payoff w also without exerting effort, the second constraint, which ensures that he is sufficiently incentivized to actually exert effort, becomes now

$$\mu \int_{s^*}^1 [b - w\psi [1 - q(s)]] f(s) ds \geq c_S. \quad (10)$$

Again, by optimality (10) will bind with equality. (We omit a formal proof as the argument is analogous to that in Proposition 1.) Substituting this into (9), we obtain the following result.¹²

¹¹More explicitly, we have that $q(s) := \frac{\pi g_h(s)}{\pi g_h(s) + (1-\pi)g_l(s)}$ and $f(s) := \pi g_h(s) + (1-\pi)g_l(s)$.

¹²With perfect claw-back, $\alpha = 1$, the agent's rent would be equal to

$$\left(\frac{c_S}{\mu\psi} \right) \frac{1 - q[1 - p(s^*)]}{\int_{s^*}^1 [p(s) - p(s^*)] f(s) ds}.$$

Proposition 3. *Suppose the firm can influence the agent's advice more gradually, namely by ensuring that the agent only sells in case of observing a sufficiently high signal $s \geq s^*$. Then the firm's total costs from ensuring that the agent both exerts sales effort and indeed follows this rule are now*

$$C = c_S \left[1 + \frac{1}{\psi} \frac{1}{\mu} \frac{1}{\int_{s^*}^1 [q(s) - q(s^*)] f(s) ds} \right]. \quad (11)$$

Expression (11) mirrors expression (8) from our basic model. Again, the double-task problem creates additional costs for the firm. Furthermore, in analogy to Corollary 1, we could again observe that these costs are strictly lower the more precise the information is as ψ is higher or the more likely it is that a sales effort leads to an interested customer as μ is higher. In addition, inspection of (11) immediately reveals that the firm's costs are strictly higher the higher the threshold s^* that it wants to implement, i.e., the more discriminatory it wants the agent to become regarding the set of types to which the good or service shall be sold. Moreover, a key observation for our analysis is that the marginal costs of increasing s^* for the firm are strictly higher the higher the agent's costs of effort c_S . Likewise, as c_S increases the effect on the firm's costs will be higher the higher is s^* .

Corollary 2. *We have for the firm's (wage) costs that $dC/ds^* > 0$ and that $d^2C/(dc_S ds^*) > 0$.*

Sales incentives and the potential for misselling

The expected time and effort it takes for an agent until he realizes a successful sale should, amongst other things, depend on the degree of competition prevailing in the market. In the most simple case, we could imagine that by contacting a customer the agent makes him aware of the firm's product. If there are more firms in the market and if these firms employ a very active salesforce, then it is more likely both that a given customer may already have bought a substitute or that he is currently already deliberating one or more alternative offers. In either case, the likelihood with which, say, an agent is successful when (cold-)calling a customer should be lower.

Intuitively, in this case the firm would have to increase the agent's incentives. Moreover, for a firm there could be also other reasons for why it has to or wants to provide higher sales incentives. For instance, a firm may want to enter a new market in which it is little known or there may be additional benefits associated with gaining market share more rapidly. Without having to spell out precisely how by exerting more or less effort an agent may affect (more

gradually) the likelihood of making a sale, note that the agent's marginal incentives are, for given s^* , simply obtained from¹³

$$\Delta := w\psi \int_{s^*}^1 [q(s) - q(s^*)] f(s) ds, \quad (12)$$

which, after substitution from (9), equals the agent's additional expected pay after finding a customer who is willing to buy. From Proposition 3 and Corollary 2 we thus have that if the firm wants to provide the agent with higher incentives by scaling up Δ , then if the firm still wants to implement the same threshold s^* this will come at additional costs. Moreover, we also know that the higher Δ , the larger are the *incremental* costs associated with pushing up the threshold s^* . From this we thus have the following general insight from our model. As the firm wants to scale up its agents' sales and marketing incentives, then at the same time the firm should find it optimal to lower s^* , thereby inducing its agents to recommend a purchase more often than previously.

Clearly, the extent to which such a shift in s^* occurs depends crucially on the associated costs. As noted previously, the firm may be worried about a loss in reputation or the likelihood from future prosecution for misselling. To make this more precise, we now suggest rather parsimoniously that if a customer ends up with utility u_l at a future point of time, then this will impose (in expected terms) a penalty ρ on the firm. The imposed penalty ρ lowers the firm's net proceeds when selling to a low-type customer. Denote now by m_h and m_l the respective margins (gross of the employee's compensation and, in the case of $\theta = l$, the penalty ρ) from selling to types $\theta = l, q$.¹⁴ After substituting the binding constraints, we then have from Proposition 3 that the firm's expected profits are now equal to

$$\Pi = \mu \int_{s^*}^1 [q(s)m_h + [1 - q(s)](m_l - \rho)] f(s) ds - C. \quad (13)$$

If we again capture incentives more generally by Δ as in (12), we then have that

$$\begin{aligned} \frac{d\Pi}{ds^*} &= -f(s^*)\mu [q(s^*)m_h + [1 - q(s^*)](m_l - \rho)] \\ &\quad - \Delta \frac{d}{ds^*} \left(\frac{1}{\int_{s^*}^1 [q(s) - q(s^*)] f(s) ds} \right). \end{aligned} \quad (14)$$

¹³However, for one of our further results we will spell out such a model in Appendix B.

¹⁴Differential margins $m_h > m_l$ may be natural in many settings. For instance, borrowers may pay back loans with different probabilities. Likewise, once a customer finds out that a particular investment or savings scheme is not fully suitable, he is more likely to cancel it, thereby reducing the overall expected margin for the firm.

The first observation is now that at an interior solution with $d\Pi/ds^* = 0$ we have that

$$q(s^*)m_h + [1 - q(s^*)](m_l - \rho) < 0. \quad (15)$$

That is, at s^* , where the agent is just indifferent between making a sale or not, the firm would strictly prefer that *no* sale is made. This is intuitive as we know that a reduction of s^* allows the firm to save rents from an *ex-ante* perspective. After re-arranging $d\Pi/ds^* = 0$, we have next that

$$\begin{aligned} & \rho - m_l \\ = & \frac{1}{1 - q(s^*)} \left[\frac{1}{f(s^*)} \frac{\Delta}{\mu} \frac{d}{ds^*} \left(\frac{1}{\int_{s^*}^1 [q(s) - q(s^*)] f(s) ds} \right) - q(s^*)m_h \right]. \end{aligned}$$

As the right-hand side of this expression is strictly increasing in Δ , we have the following result.

Proposition 4. *Suppose a policymaker, who can choose the penalty ρ , wants to ensure that the firm's agents do not advise a sale unless $s \geq s^*$ for a given threshold s^* . Then if it is optimal for the firm to increase the incentives for sales, as expressed by a higher Δ , the regulator must likewise increase the expected penalty ρ .*

Our simple model raises the obvious question of what would constitute the *optimal* policy, as expressed by the choice of ρ . Looking only at an individual sale, if we assume for simplicity that the firm has zero costs of serving a customer irrespective of his type, then welfare would be maximized if a threshold s^{**} satisfying $q(s^{**})u_h + [1 - q(s^{**})]u_l = 0$ was applied.¹⁵

If at least a fraction of ρ resulted in a deadweight loss, however, then the policy maker would, intuitively, induce a strictly lower threshold than s^{**} , thereby effectively accepting a certain level of misselling in equilibrium. Still a further reduction of s^* would be optimal if also some of the *additional* (compliance) costs to the firm, namely $C - c_S$, would come as a deadweight loss, though in our current model this represents simply a transfer to the agent.¹⁶ As we have seen, these additional costs are strictly higher the more “stringent” is the firm’s sales policy, as expressed by higher threshold s^* .

¹⁵Note that we assume throughout the paper that the respective utility u_θ are net of a customer’s alternative option. Also, from the perspective of consumer surplus alone, as we need to subtract a price the optimally implemented threshold would be strictly higher.

¹⁶See, however, the extension in Section 4, where the firm incurs costs of monitoring the agent.

Finally, our model tells us that the choice of ρ has also repercussions for the sales incentives that the firm will provide to its agents and, consequently, for the sales effort that they exert in equilibrium. To see this more clearly, it is useful to restate our previous results as follows.

Corollary 3. *Achieving a given level of sales incentives, as expressed by Δ , is strictly more expensive if the firm wants at the same time to implement a higher threshold s^* .*

The higher s^* , which as we have seen can be achieved by the policymaker through choosing a higher ρ , the more expensive it becomes for the firm to provide still the same level of sales incentives. How this should affect the optimal choice of ρ depends, however, on whether sales effort is considered to be (purely) wasteful, aimed primarily at stealing away business from competitors with equally good products or services, or whether it is necessary to create consumers' awareness of the respective product or service in the first place.

3.2 Misselling with Endogenous Prices

For our previous analysis we kept constant the firm's margins when we changed the firm's internal (compensation) policy. There are two justifications for it. First, if the firm's choice of (w, b) is not observable, which is what we assume, then for its marginal incentives, as expressed by the derivative (14), we can indeed ignore any repercussions via a possible adjustment of prices and margins. Second, depending on customers' beliefs, we could also suppose that the firm may refrain from adjusting the price as this could be interpreted by customers as, say, a sign of an (adverse) change in the firm's selling policy and thus the quality of advice.

To show the robustness of our results and to develop additional implications, though, we now endogenize also the firm's margins. We do this for the most simple case where these are independent of the customer's type. Moreover, though this is only for convenience, we abstract from any direct costs of sales such that $m_\theta = p$, where p is the prevailing price of the good or service. Moreover, we focus on the case where the firm wants its agent to provide advice and, consequently, would not want to invoke a disclaimer stating that its salesforce cannot be held legally responsible for any advice. While the latter strategy would potentially avoid any penalty, it would also lower the maximum feasible price to¹⁷

$$p = u_\emptyset = \int_0^1 [q(s)u_h + [1 - q(s)]u_l] f(s)ds = \pi u_h + (1 - \pi)u_l.$$

¹⁷See, however, Section 4 for a consideration of such a strategy.

Moreover, we now stipulate first that the whole penalty ρ represents a pure deadweight loss. In particular, this implies that it does not enter into customer's expected utility when purchasing the good. (See, however, the discussion below.)

We now consider the following game. First, the firm announces a price p . Second, the firm chooses the compensation structure (w, b) , which is not observed by outsiders. Finally, the firm's agent contacts a customer and potentially realizes a sale. By choosing this particular timing, namely that p is set before (w, b) , we overcome the multiplicity problem that would otherwise be generated by a game of signaling.

It is next convenient to repeat the key equations from the previous Section, where we now substitute $m_\theta = p$. We then have for the firm's profits

$$\begin{aligned}\Pi &= \mu \int_{s^*}^1 [p - [1 - q(s)]\rho] f(s) ds - C, \\ C &= \frac{\Delta}{\int_{s^*}^1 [q(s) - q(s^*)] f(s) ds}.\end{aligned}$$

We assume that, holding p fixed, the program to maximize Π is strictly concave and that the unique solution satisfies $s^* > 0$. It is then characterized by the requirement that

$$-f(s^*)\mu [p - [1 - q(s^*)]\rho] - \Delta \frac{d}{ds^*} \left(\frac{1}{\int_{s^*}^1 [q(s) - q(s^*)] f(s) ds} \right) = 0. \quad (16)$$

Importantly, note here that for (16) we left the price p unchanged as we vary s^* . Implicit differentiation of (16) yields then a strictly decreasing function $\widehat{s}^*(p)$. Next, to ensure that the customer is willing to buy, it must hold that

$$p \leq \frac{\int_{s^*}^1 [q(s)u_h + [1 - q(s)]u_l] f(s) ds}{1 - F(s^*)}, \quad (17)$$

where from the customer's rational expectations we must substitute $s^* = \widehat{s}^*(p)$. By optimality (17) must be satisfied just with equality, which gives rise to a continuous and strictly increasing function $\widehat{p}(s^*)$.

Proposition 5. *With an endogenous price, we have the following results.*

- i) An increase in Δ leads to a strict reduction in s^* and p .*
- ii) An increase in ρ leads to a strict increase in s^* and p .*
- iii) To ensure that the same threshold s^* is chosen, the higher Δ the higher must also be ρ .*

Proof. We argue first more formally why (17) must bind in equilibrium. For this note first that Π is strictly increasing in p , where we make use of the envelope theorem. Using continuity, (17)

must thus bind once we substitute $s^* = \widehat{s}^*(p)$. Summing up, we have a unique equilibrium where the price p and the resulting cutoff s^* are jointly determined through the two requirements that $s^* = \widehat{s}^*(p)$ and $p = \widehat{p}(s^*)$.

For assertion i) we can write explicitly $\widehat{s}^*(p, \Delta)$, which has the partial derivative $\widehat{s}_{\Delta}^* < 0$. The assertion follows then as also $\widehat{s}_p^* < 0$ and $\widehat{p}_{s^*} > 0$. For assertion ii) we write $\widehat{s}^*(p, \rho)$ with $\widehat{s}_{\rho}^* > 0$. The assertion follows then again from $\widehat{s}_p^* < 0$ and $\widehat{p}_{s^*} > 0$. Finally, assertion iii) follows immediately from assertions i)-ii). **Q.E.D.**

Note that we assumed so far that the choice of (w, p) is unobservable. Suppose next, instead, that this could be credibly communicated. For instance, it could be part of a firm's regulatory requirements to lay open the commissions that it pays to its agents. In fact, the UK's financial service regulator has imposed such transparency requirements for particular products that are sold through advisors.

It is immediate that this commitment will strictly benefit the firm. Note next that with this change the first-order condition of maximizing Π w.r.t. s^* becomes now¹⁸

$$\begin{aligned} & \mu \frac{d\widehat{p}(s^*)}{ds^*} \int_{s^*}^1 [\widehat{p} - [1 - q(s)]\rho] f(s) ds \\ & - f(s^*)\mu [\widehat{p} - [1 - q(s^*)]\rho] - \Delta \frac{d}{ds^*} \left(\frac{1}{\int_{s^*}^1 [q(s) - q(s^*)] f(s) ds} \right) \\ & = 0. \end{aligned} \tag{18}$$

Hence, compared to the first-order condition in (16) we have now an additional (and strictly positive) term in the first-line of (18), which represents the immediate feedback that the adjustment of (w, p) and thus the implementation of a different s^* have on the price \widehat{p} . As in the equilibrium without transparency the respective condition (16) must also hold at $p = \widehat{p}(s^*)$, we have the following result.

Proposition 6. *If the firm's compensation policy can be credibly made transparent, then for given ρ this leads to a strictly higher choice of s^* and thus also p .*

In the remainder of the Section, we discuss the implications of our specification that all of ρ represented a deadweight loss. More precisely, what was important for our analysis is that

¹⁸Note that to make clear the fact that also the price reacts to a change in the contract and thus to a change in the thereby implemented threshold, we have already substituted $p = \widehat{p}(s^*)$.

that ρ does not directly affect a customer's expected payoff. Assume now instead that a fraction $0 \leq \beta \leq 1$ of ρ can be appropriated by a low-type customer through compensation. This changes the requirement (17) into

$$p \leq \frac{\int_{s^*}^1 [q(s)u_h + [1 - q(s)](u_l + \beta\rho)] f(s) ds}{1 - F(s^*)}. \quad (19)$$

(Note that we assume now again that the compensation structure, or more importantly any change to it, is not observable.) We have the following result.

Proposition 7. *With endogenous prices, if a low-type customer can appropriate the fraction β of the penalty ρ through compensation, then this leads, all else unchanged, to a strictly lower s^* and a strictly higher price.*

Proof. By optimality (19) must again be binding (cf. Proposition 5). The resulting mapping $\widehat{p}(s^*)$ is also shifted upwards the higher β . Hence, the fixed-point (s^*, p) must lie at a higher value of p given that β does not affect $\widehat{s}^*(p)$. The fact that s^* is then strictly lower follows simply from strict monotonicity of $\widehat{s}^*(p)$. **Q.E.D.**

Note that the interaction in Proposition 7 would be absent if the price was not endogenized. This follows as β does not directly affect the firm's own trade-off, holding p constant. As in equilibrium, however, for given s^* the feasible price p is strictly higher the higher β , the firm has higher incentives to expand sales by lowering s^* . As only a smaller fraction of ρ now represents a deadweight loss, though, also the threshold s^* that would maximize welfare should be lower. The extent to which the *efficient* threshold changes would depend, however, also on how the firm would optimally adjust its sales incentives, Δ , and what effect this would in turn have on welfare (cf. the discussion at the end of Section 3.1).

3.3 Misselling under Different Sales Processes

Separated advice

So far our focus was squarely on a situation where a single agent performs two roles at the same time, namely that of marketing the product through, for instance, calling up potential clients and that of explaining the product and advising clients on its suitability. Depending on the nature of the product or service and a firm's internal organization, these tasks may, however, also be split. In the first part of this Section we analyze the implications if this is the case.

For an illustration, consider an “old-fashioned” bank where the staff basically waits for clients to come through the door.¹⁹ Clients may in turn have been brought in by the general marketing efforts of the firm. Even if, so as to stay in the confines of our model, we supposed that the marketing campaign is devised by a single agent at private disutility c_S , then it is clear that given the discreteness of the effort problem the respective agent would not realize a rent. More precisely, if the firm implements a sales policy described by the threshold s^* , then the agent could simply be compensated with a zero baseline wage together with a bonus equal to $c_S/(\mu[1 - F(s^*)])$.²⁰ That is, with this organization the first task of marketing the product comes then only at costs c_S .

To deal with the second task, namely that of advising incoming customers whether to purchase the advertised good or service, we specify that gathering the relevant information comes more generally at costs $c_A \geq 0$. Precisely, c_A is the agent’s private disutility from familiarizing himself with the circumstances of a particular customer, resulting in the realization of an informative signal s . Note that for $c_A > 0$ we then have *two* conditions that must be satisfied to ensure that the customer is properly advised. First, the agent has to exert additional effort at costs c_A . Second, the agent must then still provide the right advice, thereby forsaking a sale after observing $s < s^*$.

The second condition is again given by (9), which pins down s^* for a given choice of w and b . To formulate the first condition, note that the agent’s payoff from indiscriminately advising a customer to buy, given that the agent does not even observe s , is simply

$$\int_0^1 [b + w [1 - \psi [1 - q(s)]]] f(s) ds.$$

In contrast, after exerting effort at costs c_A , the agent will optimally advise the customer not to purchase if $s < s^*$, in which case he earns w . Comparing the respective payoffs and rearranging terms, we obtain the condition that²¹

¹⁹In fact, in many countries it is also not legal to undertake prospecting (or cold-calling) for certain types of retail financial products such as personal loans.

²⁰Note here that by specifying that agents have a market wage of zero we abstract from immediate efficiency gains that would arise from employing only a single agent.

²¹Note that this implies realistically that the firm cannot distinguish whether no sale was made simply because no customer approached the agent, say the staff in a bank’s branch, or whether the agent advised against a purchase.

$$\int_0^{s^*} [w\psi [1 - q(s)] - b] f(s)ds \geq c_A. \quad (20)$$

Note that for $c_A = 0$ condition (20) simplifies to (9).²² Intuitively, condition (20) is easier to satisfy if it is less costly for the agent to acquire the information that is necessary to provide advice.

For a given threshold s^* , the firm's objective is again to minimize wage costs. Intuitively, the (advising) agent once again realizes a rent equal to w , which comes on top of the respective costs of effort. We have the following result.

Proposition 8. *Suppose a given product is sold in a way that the marketing part of bringing in customers and that of giving advice are separated. Then for given s^* the firm incurs total costs of*

$$C = c_S + c_A \left(\mu + \frac{1}{\psi \int_0^{s^*} [q(s^*) - q(s)] f(s)ds} \right). \quad (21)$$

Proof. Substituting b from (9) into (20), which binds by optimality, we have that

$$w = \frac{1}{\psi \int_0^{s^*} [q(s^*) - q(s)] f(s)ds} c_A. \quad (22)$$

The agent's rent w from (22) has then to be added to the expected effort costs to obtain $C = w + c_S + \mu c_A$, which finally becomes (21). **Q.E.D.**

We ask now how the firm's different organization would affect its optimal choice of s^* . Observe once again that the choice of s^* does not affect the costs of the first activity. (More precisely, if this is delegated to an agent then the agent is never paid a rent.) Substituting now the costs C from (21) into the profit function (13), which incorporates the penalties ρ , we then have that

$$\begin{aligned} \frac{d\Pi}{ds^*} &= -f(s^*)\mu [q(s^*)m_h + [1 - q(s^*)](m_l - \rho)] \\ &\quad - \frac{c_A}{\psi} \frac{d}{ds^*} \left(\frac{1}{\int_0^{s^*} [q(s^*) - q(s)] f(s)ds} \right), \end{aligned}$$

where the last term is now strictly positive. Hence, in sharp contrast to our previous results, namely condition (15), we now have that

$$q(s^*)m_h + [1 - q(s^*)](m_l - \rho) > 0. \quad (23)$$

²²It turns out that we can omit the possibility that the agent does not incur costs c_A and simply advises the customer *not* to buy.

That is, under task separation we have that under the optimal choice of s^* the firm would now strictly *prefer* to sell at the threshold s^* . (Note that by definition the agent is just indifferent at s^* .) In contrast, if the agent performs both tasks we had in (15) the opposite sign.²³ This observation implies now immediately the following results.²⁴

Proposition 9. *For given ρ , the firm optimally implements a strictly higher threshold s^* if the task of advising customers is allocated separately to an agent compared to the case where the agent would also perform the task of contacting customers. Moreover, to obtain a given s^* a policymaker would thus have to choose a lower value ρ .*

Incidentally, note also that in the current case the firm should find it optimal to spend more on marketing and sales or, likewise, to provide the respective agents with higher incentives. This follows as this no longer interferes with the incentives of its other agents, namely its advisors. (Recall that if one agent is also responsible for the first task, then holding s^* constant this creates additional costs through higher rents.)

Optimal task allocation

Economies of scope as well as the nature of a particular product may make a particular choice of the sales process optimal. In the rest of this Section we abstract from this and suppose, instead, the firm could choose to allocate the two tasks to one or two agents without thereby forsaking efficiencies or incurring additional costs.

We first extend our basic analysis to the case where more generally $c_A \geq 0$ holds. To induce the right advice, the two conditions (9) and (20) must again hold. Moreover, taking into account the agent's additional costs c_A , condition (2), which ensures that the agent exerts sales effort at private costs c_S in the first place, becomes now

$$\mu \left[\int_{s^*}^1 [b - w\psi [1 - q(s)]] f(s) ds - c_A \right] \geq c_S, \quad (24)$$

which is clearly harder to satisfy the higher c_A .

A comparison of the two cases with and without task separation yields the following result.

²³Note again from our previous remarks that these observations also hold if we endogenize the respective margins m_θ through allowing the firm to extract the expected consumer rent.

²⁴Admittedly, we have so far not solved the program from our initial problem in case $c_A > 0$. See, however, the proof of Proposition 7.

Proposition 10. *Holding s^* fixed, it is always (weakly) more profitable for the firm to allocate the two tasks of contacting customers and of providing advice to two different agents. This holds strictly whenever c_S/c_A is not too small.*

Proof. See Appendix.

The driving force of this result is germane to the one driving separation of tasks in other multitask models. In our model, a single agent must be compensated in the same way when failing to sell to a customer, regardless of whether this failure is due to the fact that there was no customer or that there was a customer with a low type. By separating the two tasks, each agent instead can be compensated for success in the separate task performed. Following a similar logic, Dewatripont and Tirole (1994) show that separation improves incentive provision to advocates whose compensation when providing conflicting pieces of evidence must be the same as when producing no evidence.

Inspecting the proof of Proposition 10, we have next the following result.

Corollary 4. *The benefits from separating the two tasks are strictly higher if agents must be provided with higher incentives for the first task of contacting customers, given that c_S is higher and μ lower.*

Corollary 4 brings out more clearly why task separation is optimal in the first place. As we have learnt already from our most basic model in Section 2, the fact that the agent must be compensated for exerting sales effort, which is done through a bonus if a sale is realized, subsequently biases him towards indiscriminately advising customers to buy. Unless the agent must already realize a sufficiently high rent from his second task of obtaining information at costs $c_A > 0$, this makes it necessary for the firm to increase the agent's rent and thereby its wage costs if it still wants to implement the same threshold s^* .

To conclude this Section, it is illustrative to compare Proposition 10 to the outcome that is obtained if the firm can separately remunerate both tasks even if they are performed by a single agent. For instance, the firm may be able to verify whether or not the agent has contacted a customer. We then have the following result.

Proposition 11. *Suppose now, in contrast to our previous assumptions, that it is possible to separately remunerate the two tasks even if they are allocated to the same agent. Then it is always strictly profitable to allocate both tasks to a single agent.*

Proof. See Appendix.

Proposition 11 mirrors similar findings in the literature. Several papers have found how limited liability can create such “economies of scope” in task allocation purely through incentive considerations (e.g., Baron and Besanko (1992), Dana (1993), or more recently Laux (2001)).²⁵

3.4 (In-)Feasibility of Avoiding Misselling

We return now to the case where it is too costly or infeasible to separate the two tasks. Our contractual specifications so far allowed, albeit at potentially high costs, to implement any desired threshold s^* . As can be easily seen, as s^* becomes increasingly large, thereby narrowing the set of customers to which the firm wants to sell, the firm’s costs and thus the agent’s rent increase beyond any bound (cf. expression (11)). Besides being potentially no longer profitable, such contracts may also not be feasible. A too generous “base wage” w , which an agent receives even without making any sale, may backfire by attracting agents that may have little or no skill and expertise in acquiring customers. In the extreme case, such agents would only apply to earn the sure rent without being able to contribute to the seller’s business.²⁶

Generally, we might suggest that these considerations impose some upper bound on w . In what follows, we simplify the analysis by specifying that this bound is zero, though our qualitative insights apply for any such bound. In this case, the firm has only one instrument for compensation, namely the bonus b . In a one-shot setting, it is then clearly no longer feasible to provide adequate incentives along both tasks of the agent. As has long been recognized by the efficiency-wage literature, however, in a repeated context the threat to lose *future* rents can provide powerful incentives.

We consider the following incentive scheme. As before, the agent is paid b if a sale was made in a given period. In addition, he is now fired with some probability v if after a sale negative information was revealed, while he continues to be employed otherwise. Denote $\hat{\psi} := v\psi$. Also, we denote by δ the agent’s discount factor.

²⁵One difference is that in these papers tasks are typically performed simultaneously, while even with task separation a rent must be paid to both agents. Intuitively, in our setting the sales effort will now no longer be a source of rent to either of the agents given that it is discrete and that its outcome is now perfectly verifiable.

²⁶In the corporate finance literature, these are often called “fly-by-night” contractors. The existence of such unserious or fraudulent agents is there typically invoked to rule out such flat (up-front) payments.

Given stationarity of the incentive scheme, we can denote the agent's expected rent at the beginning of any period by R . If the agent recommends a customer to purchase the good, then his expected payoff for given s equals $b + \left[1 - \widehat{\psi} [1 - q(s)]\right] R$. If there is no sale, instead, then the agent just realizes δR . Consequently, the threshold s^* now satisfies

$$b = \widehat{\psi} [1 - q(s^*)] \delta R. \quad (25)$$

Note next that in case the agent applies the threshold s^* , we have also that²⁷

$$R = \mu \int_{s^*}^1 \left[b + \left[1 - \widehat{\psi} [1 - q(s)]\right] \delta R \right] f(s) ds - c_S. \quad (26)$$

After substituting from (25), we have finally from (26) that

$$R = \frac{c_S}{\delta \mu \int_{s^*}^1 \left[1 + \widehat{\psi} [q(s) - q(s^*)] \right] f(s) ds - 1}. \quad (27)$$

Irrespective of its own discount factor, the firm wants to minimize the agent's rent R . (Note that the firm's wage must also compensate the agent for c_S .) Inspection of (27) shows that this is intuitively the case by setting $v = 1$ and thus choosing the maximum feasible value $\widehat{\psi} = \psi$.²⁸

Inspection of (27) shows next that all our previous comparative results continue to hold, namely that the firm's additional costs are strictly higher the higher s^* and c_S and the lower μ . Intuitively, the additional costs are now also higher the more impatient the agent is given that δ is lower.

More interestingly, we now have the following additional result.

Proposition 12. *If the firm cannot reward an agent above his market wage in case he made no sale and can thus only punish him by severing the relationship, then a cutoff s^* can only be implemented in case it satisfies*

$$\int_{s^*}^1 [1 + \psi [q(s) - q(s^*)]] f(s) ds > \frac{1}{\delta \mu}. \quad (28)$$

Recall now our previous way of endogenizing the choice of s^* through the imposition of some expected penalty ρ in case of a sale to a low-type customer. The highest $s^* < 1$ that is now

²⁷We set again $c_A = 0$.

²⁸This presumes, of course, that replacing the agent comes at no additional costs to the firm.

feasible is determined by making (28) just binding. Moreover, as the firm moves closer to this boundary, the rent that it must leave the agent, R , increases steeply. (For the implications for costs, note that in case the firm uses the same discount factor we now have $C = R + c_S/(1 - \delta)$.) Even large changes in ρ , which in case a fraction of it represents deadweight loss may create sizable welfare costs, will then have only small and ultimately negligible effects on s^* .

Finally, it is also worthwhile to observe that now the agent's incentive constraint for the provision of effort is always slack. This is intuitive as the agent simply obtains a payoff of zero if he does not even try to contact a customer. One implication of this is the following. Suppose that there is some shock that reduces the expected per-period payoff for the agent, R . For instance, it could become more costly for the agent to locate a customer given that competition has increased. (That is, c_S may have gone up.) Suppose also that this shock is not too large so that, given that the constraint was previously slack, it is still optimal for the agent to exert effort. As R decreases, however, the agent will adjust s^* downwards (cf. (25)). If the firm cannot respond sufficiently fast, say as the agent is closer to the market, then in this (transitory) time we should observe more misselling than expected.

4 Further Discussion of Basic Model

4.1 Misselling and Deceptive Selling

The interpretation we chose for our model was that the agent must advise a customer on the suitability of a good or service for his particular needs. For instance, for retail investments the product may have a particular risk-return profile that is not optimal for all investors or may result in particular tax advantages, possibly at the cost of higher risks. Besides pointing out the characteristics of the good, the agent then finds out about the characteristics of the customer, possibly at additional costs c_A , and uses this information to advise on whether the customer should make the purchase (or investment).

However, our model allows for a much wider interpretation. In particular, the agent may not offer any advice at all! Instead, his task could be simply to explain a potentially complex product or service without inquiring about the customer's particular circumstances. Besides financial products, we could think of calling plans or long-term utility contracts, prescribing different prices for different times of the day and different consumption profiles. Here, a customer may

already have a contract. The task of the agent is then to both contact the customer and to convince him to switch to the new plan.²⁹

Suppose the good or service may have only two different “qualities”. (A quality level could comprise here all factors that determine the gross utility that it can deliver to any customer.) Customers have still different net valuations as they may, for instance, have different existing contracts or may have located substitutes at different prices. Suppose that there are again two types of customers. If the new good has only low quality, then the respective net valuations satisfy $\underline{u}_l^R < 0 < \underline{u}_h^R$, implying that only the high type would buy. If instead the new good has high quality, then with respective valuations $\bar{u}_\theta^R > 0$ both types would buy.

Suppose now that a firm’s good or service has only the lower of the two quality levels. (In fact, this should be the generic case as there should *always* be the possibility that the agent can pretend that the good delivers a still higher utility.) If the agent now pretends that the good or service is of the higher quality, then this increases the likelihood of a sale from π to one.³⁰ Provided that the firm wants to prevent this, while still eliciting sales effort, the contract (w, b) must again satisfy the two constraints from our basic analysis.

4.2 Costly Monitoring

We stipulated so far that with probability ψ there will be a feedback to the firm in case the good was sold to a low-type customer. We suggested that this could be due to complaints or other actions taken by dissatisfied customers such as the cancellation of a contract. In particular in regulated industries, where firms must maintain a given level of compliance and due diligence, firms may spend additional resources to monitor and review the sales process, thereby ensuring that, in the language of our basic model, sales are only made to high-type customers.³¹

²⁹Compare ‘Protecting Citizen and Consumers from Mis-selling of Fixed-Line Telecommunication Services’, Notification of Modification to a General Condition, Ofcom, 13 April 2005, <http://www.ofcom.org.uk/consult/condocs/misselling/statement.pdf>.

³⁰We omit here a full specification of the complete game, including possibly the choice of quality and, in particular, the setting of the sales price. For the analysis we may suppose that regardless of the quality of its good the firm sets the same (pooling) price.

³¹Firms may also adopt different policies depending on the particular sales channel. An example of such a monitoring process is given in a case description by the FSA, the UK’s financial services regulator (see <http://www.fsa.gov.uk/pages/Doing/Regulated/tcf/case/networks.shtml>). There, it is reported that the particular firm has the target “checking 10% of fact finds” for each appointed representative and for each adviser.

To incorporate this we stipulate now instead that generating information is costly and that the firm can choose to what extent it monitors sales. Precisely, the respective probability is given by ψ and the firm can once again detect for sure a sale to a low-type customer.³² Reviewing a sale comes, however, at costs $k > 0$.

As for Proposition 2 of our basic model, we allow again more generally that the fraction α of the initial bonus b can be clawed back. The agent's respective incentive constraints are then $\psi w \geq b(1 - \psi\alpha)$ and $\mu\pi b \geq c_S$, which correspond to equations (5) and (6) in the proof of Proposition 2. From these binding constraints, we then obtain the base wage

$$w = c_S \left(\frac{1 - \psi\alpha}{\psi} \frac{1}{\mu\pi} \right), \quad (29)$$

as in (7). The difference to Proposition 2, however, is that the firm's expected costs are

$$C = c_S + w + \mu\pi k\psi, \quad (30)$$

given that monitoring is now costly. For given b , the firm can thus ensure compliance by the agent either through an increase in ψ or an increase in w , which both are costly. The optimal combination of w and ψ minimizes the total costs (30). We have the following results.

Proposition 13. *Suppose that in our basic model ψ represents the monitoring intensity that the firm optimally chooses. If monitoring is sufficiently costly, $k > c_S (\mu\pi)^2$, we have an interior solution with $\psi = \frac{1}{\mu\pi} \sqrt{\frac{c_S}{k}}$ and $w = \sqrt{c_S k} - \alpha c_S \frac{1}{\mu\pi}$. The resulting total costs to the firm*

$$C = c_S + 2\sqrt{c_S k} - \alpha c_S \frac{1}{\mu\pi} \quad (31)$$

again satisfy $dC/dc_S > 1$.

Proof. Substituting w from (29) into (30), the firm's program is to minimize

$$c_S \left(\frac{1 - \psi\alpha}{\psi} \frac{1}{\mu\pi} \right) + \mu\pi k\psi.$$

This program is strictly convex in ψ and yields a unique solution $\psi > 0$. If

$$\frac{c_S}{k} < (\mu\pi)^2 \quad (32)$$

the solution is interior, with

$$\psi = \frac{1}{\mu\pi} \sqrt{\frac{c_S}{k}}, \quad (33)$$

³²Note that the process of monitoring is itself not verifiable.

which is thus not affected by α . Substituting (33) into (29) we have $w = \sqrt{c_S k} - \alpha c_S \frac{1}{\mu\pi}$ and thus obtain (31) from (30). Here, $dC/dc_S > 1$ holds whenever $\sqrt{k/c_S} > \alpha/(\mu\pi)$, which is implied by (32). **Q.E.D.**

5 Conclusion

This paper analyzes the plight of a firm that sells through agents, who can be either its own employees or independent intermediaries. When it comes to compliance to selling standards, the firm must make sure that these agents give customers the right advice on the suitability of the product or service sold and avoid deceptive statements that inflate the perceived value of the offering.

Our starting point is the insight that if the firm finds it profitable or necessary to incentivize its agents more to market the product by contacting customers, then this selling task interferes with the second task, namely that of providing adequate advice and avoiding deceptive claims. Aligning the agents' interests with those of the firm can then be quite costly, involving both additional rents that for the agents and additional costs related to the internal review and monitoring process that the firm puts into place to assure compliance.

A policymaker who wants to ensure that representations are not misleading and that customers make the most suitable choice must take into account the firm's internal agency problem when determining his optimal policy. In particular, if a change in the firm's (competitive) environment will possibly make it more profitable or even necessary for firms to power up the incentives of its salesforce, then a policymaker should also step up the intensity with which it monitors and prosecutes potential misselling. On the other hand, also the feedback of such policies on the incentives that a firm will optimally provide to its salesforce must be taken into account. We found here that a marginal increase in incentives becomes more expensive for the firm due to the additional compliance costs the higher the "standard" imposed through a stricter policy regime. For new products, for which customer awareness is still limited, a too strict policy of prosecuting potential misselling may then lead to an inefficiently low or slow penetration of the market, causing detriment to consumer.

We further showed how the optimal policy of prosecuting misselling may depend on both the organization of the firm's sales process, namely whether the function of advice or contract approval is performed separately, as well as on the simultaneous use of other policy instruments

such as forcing firms to make their commissions transparent. In either case, a more lenient standard of prosecuting misselling would be advisable.

Future work could embed our basic model into different contexts. Such an analysis could also throw light on the differences in both the regulatory frameworks as well as the organization of distribution across industries and countries. (The retail financial industry provides here a natural example.) The choice of policy may then dictate the optimal organizational responses, while on the other side the prevailing organization of the industry, such as the level of vertical integration, use of independent intermediaries, or also the level of competition, could in turn dictate the optimal policy response.

6 Appendix

Proof of Proposition 10. We turn first to the case in which tasks are allocated to a single agent. Clearly, either one of the two constraints (20) and (24) must be binding as, otherwise, the firm could reduce both b and w while still satisfying (9) for given s^* . Note next that (24) transforms to

$$w \geq \frac{1}{\psi} \frac{c_S/\mu + c_A[1 - F(s^*)]}{\int_{s^*}^1 [q(s) - q(s^*)] f(s) ds}, \quad (34)$$

while (20) transforms to

$$w \geq \frac{1}{\psi} \frac{c_A}{\int_0^{s^*} [q(s^*) - q(s)] f(s) ds}. \quad (35)$$

Hence, we have that (24) is binding while (20) remains slack in case the threshold for w in (34) exceeds that in (35). This is in turn the case if

$$\frac{c_S}{c_A} > \mu \left[\frac{\int_{s^*}^1 [q(s) - q(s^*)] f(s) ds}{\int_0^{s^*} [q(s^*) - q(s)] f(s) ds} - [1 - F(s^*)] \right]. \quad (36)$$

Turn now to the case with separated tasks. The analysis is the same as in Proposition 8. That is, we now have for the firm's total costs $C = w + c_S + \mu c_A$, where w is given by (22), which in turn is clearly the same as in (the binding) constraint (35). Summing up, we thus have that the total costs C are strictly higher if one agent performs both tasks in case (36) is satisfied for given s^* , while otherwise C is not affected by the separation of tasks. **Q.E.D.**

Proof of Proposition 11. Suppose first that both tasks are allocated to two different agents. Again, the compensation for the first agent costs the firm just c_S in expectation. Turning to the

second agent, the two constraints are again (9) to pin down s^* and (20) to elicit effort in case $c_A > 0$. However, the difference is now that optimally the agent only receives the compensation equal (in expected terms) to $c_A + w$ in case a customer was successfully contacted in the first place, which is now by assumption verifiable. Consequently, the firm's total compensation costs under task separation are now $C = c_S + \mu(c_A + w)$, where w is given by (22).

Suppose next that a single agent performs both tasks. Note again that the agent can now receive a separate compensation in case he did not generate a potential customer. It is intuitive that in this case the respective wage is optimally set to zero. In addition, the agent is paid according to (w, b) as previously. Note that the only difference is thus that the agent receives w *only* if a potential customer was generated. The respective constraints are then again (9) to pin down s^* and (20), while the previous constraint (24), which ensured that the agent exerted sales effort at costs c_S , becomes now instead

$$\mu \int_{s^*}^1 [b + w [1 - \psi [1 - q(s)]] - c_A] f(s) ds \geq c_S. \quad (37)$$

Again, either one of the two constraints (35) or (37) must be binding by optimality. If (37) is binding, we have that

$$w = \frac{c_S/\mu + c_A[1 - F(s^*)]}{\int_{s^*}^1 [1 + \psi [q(s) - q(s^*)]] f(s) ds}. \quad (38)$$

If this choice of w , together with the corresponding bonus b from (20), satisfies also the remaining constraint (35), then the agent does not receive *any* rent, so that $C = c_S + \mu c_A$. This is the case if w from (38) does not fall short of the threshold for w in (35), which holds in turn if

$$\frac{c_S/\mu + c_A[1 - F(s^*)]}{\int_{s^*}^1 [1 + \psi [q(s) - q(s^*)]] f(s) ds} \geq \frac{1}{\psi} \frac{c_A}{\int_0^{s^*} [q(s^*) - q(s)] f(s) ds}$$

and thus if

$$\frac{c_S}{c_A} \geq \mu \left[\frac{\int_{s^*}^1 \left[\frac{1}{\psi} + [q(s) - q(s^*)] \right] f(s) ds}{\int_0^{s^*} [q(s^*) - q(s)] f(s) ds} - [1 - F(s^*)] \right]. \quad (39)$$

Instead, if (39) does not hold, then w is characterized by the binding constraint (35), in which case the respective costs $C = \mu(c_A + w)$ are again strictly lower than those under task separation, which equal $C = \mu(c_A + w) + c_S$ for the same choice of w . **Q.E.D.**

7 References

- Baron, D.P., and Besanko, D., 1992, Information, Control, and Organizational Structure, *Journal of Economics and Management Strategy* 1, 237-275.
- Bolton, P., Freixas, X., and Shapiro, J., 2005, Conflicts of Interest, Information Provision, and Competition in Banking, forthcoming *Journal of Financial Economics*.
- Campbell, J.Y., 2006, Household Finance, *Journal of Finance* 59, 1553-1604.
- Che, Y.-K. and Spier, K.E., 2006, Strategic Judgement Proofing, Columbia and Northwestern mimeo.
- Dana, J.D., 1993, The Organization and Scope of Agents: Regulating Multiproduct Industries, *Journal of Economic Theory* 59, 288-310.
- Darby, M.R., and Karni, E., Free Competition and the Optimal Amount of Fraud, *Journal of Law and Economics* 16, 67-88.
- Demski, J.S., and Sappington, D.E.M., 1987, Delegated Expertise, *Journal of Accounting Research* 25, 68-89.
- Dewatripont, M., and Tirole, J., 1999, Advocates, *Journal of Political Economy* 107, 1-39.
- Dulleck, U., and Kerschbamer, R., 2006, On Doctors, Mechanics and Computer Specialists: The Economics of Credence Goods, *Journal of Economic Literature* 44, 5-42.
- Holmström, B. and Milgrom, P., 1991, Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design, *Journal of Law, Economics, and Organization* 7, 24-52.
- Hoyt, R.E., Dumm, R.E., and Carson, J.M., 2006, An Examination of the Role of Insurance Producers and Compensation in the Insurance Industry, mimeo.
- Lambert, R., 1986, Executive Effort and Selection of Risky Projects, *Rand Journal of Economics* 17, 77-88.
- Laux, C., 2001, Limited-Liability and Incentive Contracting with Multiple Projects, *Rand Journal of Economics* 32, 514-526.

Levitt, S.D., and Snyder, C.M., 1997, Is no News Bad News? Information Transmission and the Role of Early Warning in the Principal-Agent Model, *Rand Journal of Economics* 28, 641-661.

Lewis, T.R., and Sappington, D.E.M., 1997, Information Management in Incentive Problems, *Journal of Political Economy* 97, 796-821.

Pitchford, R., 1995, How Liable Should a Lender Be? The Case of Judgment-Proof Firms and Environmental Risk, *American Economic Review* 85, 1171-1186.

Prendergast, C., 2002, Consumers and Agency Problems, *Economic Journal* 112, C34-C51.