Competition through Commissions and Kickbacks[*]

Roman Inderst   Marco Ottaviani

January 2011

Abstract

In markets for retail financial products and health services, consumers often rely on the advice of intermediaries to decide which specialized offering best fits their needs. Product providers, in turn, compete to influence the intermediary's advice through hidden kickbacks or disclosed commissions. Motivated by the controversial role of these widespread practices, we formulate a model to analyze competition through commissions from a positive and normative standpoint. The model highlights the role of commissions in making the adviser responsive to supply-side incentives. We characterize situations when commonly adopted policies such as mandatory disclosure and caps on commissions have unintended welfare consequences.

JEL Classification: L15 (Information and Product Quality), L51 (Economics of Regulation), M52 (Compensation and Compensation Methods and Their Effects).

Product providers commonly pay commissions (often in the form of undisclosed kickbacks) to information intermediaries with the aim of influencing the intermediaries' advice to retail customers and the eventual sale of specialized offerings. This practice is controversial in a number of markets. In the health care sector, there is considerable concern that the quality of medical advice can be compromised by gifts and other inducements that physicians receive from pharmaceutical companies and other health care suppliers.[1] In insurance markets, allegations have been brought against insurance providers regarding the provision of hidden kickbacks to supposedly independent brokers.[2] In the mortgage industry, high commissions are believed to have led brokers to advise homebuyers to borrow beyond their means, fueling the current crisis.[3]

This paper investigates the allocative role of commissions and kickbacks with the objective of deriving positive and normative implications. In our model, two firms compete through commissions paid to an adviser. The adviser issues a recommendation to a customer regarding which of the two products to purchase, on the basis of private information about the match between the customer's needs and the characteristics of the products. The adviser is compensated through commissions paid by the firms. In addition, the adviser cares that the customer purchases the most suitable product, because of liability, ethical, or reputational concerns. Firms set product prices taking into account the advice customers receive.

While in traditional industrial organization models firms compete for customers who

---

[1]Inducements may take the form of consultant fees, educational grants, royalties, funding for clinical trials, or travel grants. See Michael L. Millenson (2003) for an overview of the practice of detailing drugs and physician preference items, such as hip and knee implants, cardiac stents, and mechanical devices used in spinal surgery. Professionals' self-imposed standards of disclosure to patients are often ineffective, as in a salient recent case involving orthopedic devices for hip and knee replacement, in which physician consultants allegedly received over $800 million from manufacturers between 2002 and 2006. See the agreed settlement between four leading manufacturers of orthopedic devices and the Department of Justice at http://www.usdoj.gov/.

[2]A recent high-profile case was brought by the former New York State Attorney General, Eliot Spitzer, against US insurance providers, most notably AIG. See David Cummins and Neal Doherty (2006) for a discussion and an empirical analysis of brokerage intermediation in the insurance market. See Howell E. Jackson and Laurie Burlingame (2007) and Daniel Schwarcz (2007) for a legal perspective on the pervasive use of commissions and kickbacks to compensate insurance, investment, real estate, and mortgage brokers.

[3]For reports by media, consumer groups, and policy makers, see: Liam Pleven and Susanne Craig, "Deal Fees under Fire Amid Mortgage Crisis; Guaranteed Rewards of Bankers, Middlemen Are in the Spotlight," *Wall Street Journal*, January 17, 2008; Ernst Keith, Debbie Bocian, and Wei Li (2008); and "Subprime and Predatory Mortgage Lending: New Regulatory Guidance, Current Market Conditions, and Effects on Regulated Financial Institutions: Hearing before the Subcommittee on Financial Institutions and Consumer Credit of the House Committee on Financial Services," 110th Congress, 2007. See also Basel Committee on Banking Supervision (2008) on the role of suitability in the advice leading to the sale of retail financial services.

have private information about which product best suits their needs, in our model customers are initially uninformed and must obtain this information from the adviser. Each firm is in a position to steer the adviser's recommendation by stepping up its respective commission. The adviser trades off earnings from commissions with the concern for a suitable choice by the customer. As we show, this trade-off is analogous to the one the final customer faces in Harold Hotelling's (1929) classic model of price competition with horizontally differentiated products. While in the Hotelling model the marginal customer reacts to the difference in products' prices, in our model the adviser reacts to the difference in firms' commissions. The adviser's concern for suitability in our model plays the role of unit transportation cost and, thus, measures the inverse of the adviser's sensitivity to commissions from competing firms.

When products are equally cost-efficient, we show that competition results in efficient allocation because firms have balanced incentives to steer advice, irrespective of the strength of the adviser's suitability concern. Nevertheless, commissions are higher when the adviser is less concerned about suitability, because firms then find it more effective to pay higher commissions. We also find that firms have higher incentives to step up commissions to steer advice in the absence of disclosure. Hidden kickbacks allow firms to expand market share without having to lower their price at the same time. When, instead, a firm increases a disclosed commission, it must also suffer a corresponding reduction in the amount the customer is willing to pay for the product that is then recommended more often by the adviser. Disclosure unambiguously reduces the equilibrium level of commissions.

When firms are asymmetric, mandatory disclosure of commissions can have unintended consequences. To see why, first note that a more cost-efficient firm, given its higher margin, also has a stronger incentive to steer advice and actually ends up with a higher market share. Nevertheless, the market share of this (larger) firm remains inefficiently too low. For a firm that already pays a higher commission and, thus, makes a sale with a higher probability, an additional increase in commission to attract an extra customer proves to be more expensive because it must be paid more often than by the rival firm. In close analogy to the traditional Hotelling model, we show that the market share of a more cost-efficient firm is inefficiently too low. Having a larger market share, the more cost-efficient firm finds an incremental increase in commission (like an incremental reduction in price) more costly than the less cost-efficient competitor, which has a smaller market share.

One of our key results is that, while mandatory disclosure stifles all commissions, it does so more for commissions paid by a more cost-efficient firm and, thus, results in a reduction

in this firm's market share. Disclosure then reduces efficiency when the market share of the more cost-efficient firm is already too small in the baseline non-disclosure scenario, as is surely the case when the adviser is highly concerned about suitability. Instead, when the adviser is less concerned about suitability, the market share of a more efficient firm is too large when commissions are not disclosed, so efficiency is higher with disclosure.

In an attempt to protect consumers of retail financial products, some jurisdictions mandate that brokers, financial advisers, and other intermediaries disclose to customers the commissions paid by product providers.[4] Physicians and other health care providers are commonly subject to bans (or strict caps) on the value of the gifts they are allowed to receive from providers.[5] In a drastic new regulation, the UK financial watchdog has recently imposed a ban on commissions for financial advisers.[6] Our analysis reveals that policies that chill commissions through mandatory disclosure or bans may have unintended consequences for efficiency because they inefficiently reduce the responsiveness of advice to supply-side differences. As we show, the overall impact of disclosure depends on the agent's concerns for suitability, which can also be affected by policy. We also analyze the effect of an increase in penalties for unsuitable sales due to a tightening of regulatory supervision, and we discuss the impact of caps or outright bans on commissions.

In an extension, we allow the quality of advice to be affected not only by the potential bias of the adviser, but also by the quality of the adviser's information. For example, advisers with superior training are in a position to obtain better information about suitability. We show how the quality of the adviser's information affects not only the customer's willingness to pay and the resulting product prices, but also firms' incentives to compete through commissions. In particular, when commissions are not disclosed and thus firms compete more aggressively, we find that a more informed adviser can extract a larger fraction of customers' additional benefits from the improved advice. This is because, without

---

[4]For example, in November 2008 the US Department of Housing and Urban Development strengthened the requirement imposed on third-party brokers to disclose to homeowners the payments they receive for intermediated mortgage agreements (see www.hud.gov). In the EU, the Markets in Financial Instruments Directive (MiFID) has required the disclosure of commissions on retail financial products since January 2008. In the UK, similar provisions were imposed earlier by the Financial Services Authority.

[5]For example, Minnesota's Fair Drug Marketing Law prohibits gifts over $50. Similar provisions exist in Vermont, California, Maine, West Virginia, and the District of Columbia. The *Physician Payment Sunshine Act*, currently awaiting approval by the US Congress, would require certain manufacturers of drugs and medical devices to disclose inducements given to physicians through consultant fees, educational grants, and/or travel gifts.

[6]A new regulation effective from 2012 prevents financial advisers in the UK from accepting commissions in return for recommending specific investment products. The restriction applies to the sale of investments such as pensions, annuities, and unit trusts, but not to mortgages and insurance policies. See Policy Statement 10/6 by the UK Financial Services Authority, released on March 26, 2010.

disclosure, the higher prices, which firms can charge when customers expect better advice, are passed through into higher commissions. We conclude that disclosure of commissions stifles the adviser's incentives to invest in information.

The paper proceeds by presenting our contribution to the literature in Section 1. Section 2 formulates the model. Section 3 characterizes the baseline scenario with undisclosed commissions, while Section 4 analyzes the regime with disclosure. Section 4 compares the two disclosure regimes from a welfare perspective. Sections 5 to 7 endogenize the adviser's concern for suitability through fines, the threat of losing the franchise, and professional standards. Sections 8 endogenizes information acquisition. Section 9 concludes. The main Appendix at the end of the text collects the proofs. An online Appendix in six sections discusses robustness and extensions.

1   Contribution to Literature

There is a dearth of literature on commissions and kickbacks. Bruce M. Owen (1977) discusses the role (and criticizes the regulation) of kickbacks that providers of conveyancing and title insurance services pay to real estate brokers in order to steer homebuyers. Mark V. Pauly (1979) models the role of kickbacks (or fee splitting) paid by one physician to another in return for patient referrals. Pauly posits that patients non-strategically follow the referral advice up to an exogenously given maximum level they find acceptable. If this maximum level is above the social optimal level, generalists might overrefer patients to specialists to collect the kickback. On the other hand, kickbacks can enhance efficiency because they incentivize generalists to refer patients to more cost-efficient specialists. To this early literature, we contribute an explicit analysis of the impact of commissions on the information that the adviser can credibly communicate to the customer. In our equilibrium model, the customer responds rationally to the adviser's recommendation, taking into account the impact of commissions on the information content of advice. As we show, only when commissions are not too high do they enhance efficiency by allowing customers to make decisions that better reflect supply-side cost differences.

Our baseline model contributes a tractable framework that embeds the provision of product advice by an information intermediary into Hotelling's (1929) classic model of competition between two price-setting firms. While Gary Biglaiser's (1993) middlemen and Alessandro Lizzeri's (1999) certification intermediaries provide vertical information about quality, our information intermediaries provide horizontal information about match suitability. Thus, our model relates to work by Tracy R. Lewis and David E. M. Sap-

pington (1994), Giuseppe Moscarini and Marco Ottaviani (2001), Justin P. Johnson and David P. Myatt (2006), Juan-Josè Ganuza and Josè S. Penalva (2010), and Heski Bar-Isaac, Guillermo Caruana, and Vicente Cuñat (2010) on sellers' incentives to provide information to customers. In our model, however, this information is provided by an intermediary adviser rather than directly by the sellers. It is only through commissions that the competing sellers can affect the information the adviser conveys to the customer.

While common agency models following B. Douglas Bernheim and Michael D. Whinston (1986) typically analyze settings in which multiple principals compete to influence an agent's decision, in our model the agent is a privately informed adviser who in turn communicates with a customer. In a model in which brokers steer customers to appropriate providers, Peter F. Colwell and Charles M. Kahn (2001) argue that the possibility of repeat business and existence of fixed costs may make it ex ante socially undesirable that brokers inform customers about the cost of the provider. In our model, instead, costs of providers are known; we show that disclosure of commissions increases or reduces social welfare depending on the level of the adviser's concern for suitability.

The intermediary's incentives to provide biased advice are influenced by firms' commissions, rather than being specified exogenously as in Vincent P. Crawford and Joel Sobel (1982) and in most of the theoretical literature on strategic communication. John Morgan and Phillip C. Stocken (2003) analyze communication by a sender with uncertain bias, while Ming Li and Kristóf Madarász (2008) characterize how disclosure of such an exogenous and uncertain bias affects the resulting cheap-talk equilibrium. Once the adviser's bias is endogenized through the commissions paid by competing firms, as in our model, disclosure also affects the firms' incentives to set commissions and, thus, the resulting level of the bias. While our adviser is concerned about suitability, Erik Durbin and Ganesh Iyer (2009) allow a (single) biased principal to influence the preferences of an adviser who aims to be perceived as being incorruptible. Wei Li (2010) analyzes how a biased sender may affect an intermediary agent through information provision, rather than monetary transfers as in our model.

The role of the adviser as an intermediary is also a key difference from Patrick Bolton, Xavier Freixas, and Joel Shapiro (2007). As in their model and in most of the credence goods literature, we allow product prices to be endogenous, but we depart by not allowing sellers to advise customers directly. Instead, in our model commissions steer the advice of a bottleneck agent who controls the ability of firms to access customers. Roman Inderst and Ottaviani (2009) also analyze the impact of compensation on advice, but they focus

on how a seller should optimally compensate a sales agent through a contract involving a fixed wage and a sales-related bonus pay. Because there the seller (rather than the adviser, as in the present model) is subject to liability for misselling to unsuitable customers, the agency problem becomes nontrivial only through the multi-task problem created by the need for the agent to search for customers and to advise them to purchase. In contrast, in our present setting the adviser is an independent agent for two firms and cares directly about suitability of products. This makes it costly for product providers to steer advice in their favor. The model formulated here is relevant to analyze how penalties for unsuitable advice and disclosure of commissions impact the efficiency of advice by affecting differently the incentives of competing firms.

2   Model

We consider a customer's choice of whether to purchase a single unit of one of two products, $n = A, B$. For example, the customer's choice could be between two different investment plans, one of which is more suitable than the other, based on the customer's financial condition, risk preferences, tax status, or life expectancy. In an application to health care, the two products could correspond to different medical treatments. Normalizing the customer's utility from not purchasing to zero, the valuation from purchasing depends on a binary state variable denoted by $\theta = A, B$. The customer derives utility v_h if the product matches the state and utility v_l otherwise, with $0 < v_l < v_h$. Equivalently, product A is more suitable than product B in state A, and vice versa in state B. Firms produce at respective costs c_n, and they can only reach the customer through an intermediary. Without loss of generality, we specify that firm A is weakly more cost-efficient than firm B, $c_B \geq c_A$.

Suitability.   The intermediary advises the customer on the basis of some private information about which of the two products is a better match for this particular customer. The adviser's private information is conveniently represented by a (posterior) belief that product A is more suitable, $q = \Pr(\theta = A)$, which is distributed ex ante according to the continuous distribution $G(q)$ with density $g(q) > 0$ over $q \in [0, 1]$. In Section 8 we further parametrize the quality of the adviser's information by introducing an ordering over the distribution of this posterior belief.

A key aspect of our model is that the private information about the match between the customer's particular needs and the firms' specific products is possessed by the adviser,

7

rather than by the customer (as in classic industrial organization models with a downward-sloping demand curve) or by the firms (as in signaling models).

We simplify derivations by restricting attention to distributions of posterior beliefs that are symmetric around the (common) prior belief $q = 1/2$, with $G(q) = 1 - G(1 - q)$. The restriction to symmetric distributions is customary in Hotelling models, to which we relate our setup and analysis throughout the paper. To guarantee that the firms' maximization problem is well behaved, we assume that $G(q)$ has an increasing hazard rate,

$$\frac{d}{dq} \frac{g(q)}{1 - G(q)} > 0. \tag{1}$$

Together with the symmetry of $G(q)$, condition (1) implies that the reverse hazard rate is decreasing,

$$\frac{d}{dq} \frac{g(q)}{G(q)} < 0. \tag{2}$$

The customer's expected utilities (gross of prices) for the two products are denoted by $v_A(q) := q v_h + (1 - q) v_l$ and $v_B(q) := (1 - q) v_h + q v_l$, respectively. We assume that

$$\int_0^1 v_A(q) g(q) dq = \int_0^1 v_B(q) g(q) dq = \frac{v_l + v_h}{2} < c_A, \tag{3}$$

so that advice is essential for selling either product, given our specification that $c_A \leq c_B$. This assumption guarantees that firms cannot circumvent the adviser and sell directly to the customer. To ensure that either product can be sold with advice, we further stipulate that

$$\int_{1/2}^1 v_A(q) \frac{g(q)}{1 - G(1/2)} dq = \int_0^{1/2} v_B(q) \frac{g(q)}{G(1/2)} dq > c_B. \tag{4}$$

Hence, if the adviser were to recommend the most suitable product based on the information contained in the posterior belief, the expected conditional valuation would exceed the cost of each product (given that $c_B \geq c_A$). We discuss below (and analyze in online Appendix 3) how these assumptions can be relaxed.

Concern for Suitability.   We posit that the adviser cares directly about whether the purchased product is suitable for the customer's needs. We capture the adviser's concern for suitability in a flexible way by positing that the adviser derives utility w_h when the customer purchases the more suitable product, and utility w_l otherwise. The adviser's concern for suitability is driven by the difference between these two utility levels, $w = w_h - w_l$, which plays a key role in our analysis. The adviser derives utility w_0 when the customer does not purchase any of the two products. We assume that $w_0 < w_l < w_h$ so as

to effectively restrict consideration to the choice between products A and B, irrespective of the size of equilibrium commissions. This assumption is satisfied if, for instance, the adviser when intermediating a purchase generates sufficiently large benefits from other simultaneous or subsequent transactions with the same customer.[7] We discuss below (and analyze in online Appendix 4) the case in which the adviser sometimes wants to recommend to customers not to purchase any product.

In the paper we develop three foundations for the adviser's concern for suitability:

- The adviser may be subject to a penalty following the purchase of a product that turns out to be a bad match for the customer. Then, w captures the size of this fine, which could be imposed by a professional association or a regulator.[8] This case is analyzed in detail in Section 5, where we also derive policy implications.

- The adviser's preference for suitability may arise from reputational concerns or fear of losing future business prospects when the business licence is revoked by an authority. To this end, Section 6 shows that our results carry over to a dynamic model in which the suitability concern w depends on the expected future commissions the adviser risks losing.

- The adviser may be motivated by a professional concern for the customer's well-being. In this case, we stipulate that the adviser places some weight γ on suitability: $w_l = \gamma v_l$ and $w_h = \gamma v_h$, so that $w = \gamma(v_h - v_l)$.[9] This specification is analyzed in

---

[7]In addition to the real estate broker example discussed by Owen (1977), this assumption applies to the advice provided by car dealers on the loan associated with the purchase of a vehicle (Consumer Federation of America 2004); on the controversial exemption of auto loans from the Dodd–Frank Wall Street Reform and Consumer Protection Act, see Center for Responsible Lending (2009) and "Should Auto Dealers Avoid New Regulation?" *Time Magazine*, May 19, 2010.

[8]For instance, occupational licensing procedures in various US states require mortgage brokers to maintain a minimum net worth or to post a surety bond (cf. Cynthia Pahl, 2007). In practice, surety bonds are typically posted through third parties. While these third parties are the first to be liable, they are then compelled by regulation to seek redress from the broker.

[9]For example, doctors care about the effect of treatment on the well-being of their patients, a standard assumption in health economics models (cf. Thomas G. McGuire, 2000). The problem has long been recognized by the American Medical Association (1948): "The pride of medicine as a profession has always been its freedom from the taint of barter and trade in the sick patient. . . . Nevertheless, the charge is made that some physicians have forgotten the ethical principles that prevail in the relationship between doctor and patient and have selected the surgeon willing to make the greatest division of fees rather than the one best suited to perform the operation. Ophthalmologists have sent the patient for lenses to opticians who returned a proportion of the fee rather than to the optician who rendered the highest quality of optical service. Occasionally orthopedic surgeons and others who utilize the work of the maker of braces, splints and elastic bandages have been willing to accept commissions from such manufacturers . . . From time to time criticism has been leveled against pharmacists who have offered commissions to physicians on the prescriptions sent to them and to the physicians who have accepted such commissions."

Section 7.

Timing. To influence the adviser, at period $t = 1$ firms simultaneously set their respective commissions (or fees) f_n to be paid to the adviser conditional on the sale of their product. The adviser has no wealth and, thus, cannot pay firms an upfront fee for carrying their products. We further restrict firms from demanding a payment from the adviser when their product is not sold. (Such payments are rarely observed in practice, and would not be feasible in a variant of the model in which there is positive probability that no customer arrives to the market and the bilateral contracts between each firm and the adviser cannot be conditional on whether a lack of sale is due to the absence of a customer or to the sale of a competitor's product.) At period $t = 2$, each firm sets the respective product price, p_n. Given our interest in situations in which private contracting through warranties fail, we rule out payments from or to customers that are contingent on the realized utility, v_l or v_h. At period $t = 3$, the adviser communicates to the customer by sending a message, $m = A, B$, to the customer, who then makes the final purchase decision at period $t = 4$. All payoffs are realized after the final period, $t = 4$. We abstract from discounting and risk considerations by assuming that all parties are risk neutral.

Throughout the paper, we compare two disclosure regimes: the baseline scenario in which commissions are not observed by the customer (and then act as hidden kickbacks), and the policy scenario with disclosed commissions. As we argue below, without policy intervention, firms should find it difficult to commit to disclosing all incentives provided to the adviser.

3   Steering Advice through Kickbacks

Definition of Equilibrium.   In the baseline model commissions are not disclosed. Our solution concept is perfect Bayesian equilibrium with the following restrictions. We focus on equilibria in which only pure strategies are played and in which advice is informative at $t = 3$. Such equilibria always exist, even though off-equilibrium commissions and prices may be such that an informative outcome does not exist and the customer thus does not learn from the adviser's recommendation. Note further that from restrictions (3) and (4) sales take place only in the equilibria we consider, because advice is necessary to generate gains from trade.

Furthermore, we specify *passive beliefs* according to which customers do not react to observed prices by changing their expectations about firms' unobserved commissions. Fol-

lowing Hart and Tirole (1990), this restriction to passive beliefs is frequently invoked in games of vertical contracting between a supplier and retailers who do not observe each others' contracts. Online Appendix 1 shows that our results remain valid when customers expect firms to change commissions optimally in $t = 1$ in anticipation of their own subsequent price deviation in $t = 2$, in the spirit of McAfee and Schwartz's (1993) wary beliefs. It is convenient to stipulate passive beliefs about commissions also with respect to the adviser's recommendation in $t = 3$, so that in any pure-strategy equilibrium the customer holds (point) beliefs $\widehat{f}_n$ about the respective commissions.

Advice. Given that both commissions and prices are endogenous in our model, we argue below that both products are sold with positive probability in an equilibrium in which advice is informative. Further, from $w_0 < w_l$ we can restrict consideration to only two messages for the adviser, which correspond to the two products, A and B. Ignoring the payoff-equivalent outcome in which the messages are swapped, in equilibrium the adviser's recommendation is followed by the customer. Then, the adviser's expected payoff from recommending product A equals $f_A + qw_h + (1 - q)w_l$ and from recommending product B equals $f_B + (1 - q)w_h + qw_l$. When both products are recommended with positive probability, the adviser recommends A rather than B when $q \geq q^*$, with the cutoff given by

$$q^* = \frac{1}{2} - \frac{f_A - f_B}{2w}. \tag{5}$$

Specifying that the adviser prefers to recommend A in case of indifference is clearly inconsequential given that this is a zero-probability event. For ease of exposition we define $q^* = 0$ in case $f_A \geq f_B + w$ and $q^* = 1$ in case $f_B \geq f_A + w$, even though these cases do not arise in equilibrium.

Hotelling Comparison. Expression (5) mirrors the derivation of the critical customer type in a model of price competition à la Hotelling in which the customer privately observes directly a signal about match quality, as in Moscarini and Ottaviani (2001). A customer who privately observes q directly is indifferent between purchasing from either firm when $v_A(\widetilde{q}) - p_A = v_B(\widetilde{q}) - p_B$, so that the marginal customer type is $\widetilde{q} = 1/2 - (p_B - p_A) / [2 (v_h - v_l)]$. In this version of the classic Hotelling model, the responsiveness of the marginal customer type depends on the importance of match quality for the customer's payoff, as measured by $v_h - v_l$, which can also be interpreted as transportation cost for each unit of belief travelled. According to (5), in our model the adviser's concern

11

for suitability, w, plays the role of unit transportation cost. While in Hotelling's model the marginal customer type reacts to the difference in products' prices, in our model the adviser steers customers to the respective products and reacts to the difference in firms' commissions. This comparison is further developed throughout the paper.

Price Setting. In a pure-strategy equilibrium, the customer rationally interprets the information content of advice by using the adviser's *expected* cutoff, denoted by $\widehat{q}^*$, which is obtained by plugging the *expected* commissions $\widehat{f}_n$ into (5). Again, if substitution of $\widehat{f}_n$ into (5) does not lead to an interior threshold, we set $\widehat{q}^* = 0$ or $\widehat{q}^* = 1$, respectively.

When choosing prices p_n in $t = 2$, firms have to take into account these expectations, because they determine the customer's willingness to pay for their products. For given expected cutoff, $\widehat{q}^*$, the customer's *conditional* expected valuation for each product is given by

$$
\begin{aligned}
P_A(\widehat{q}^*) &= \int_{\widehat{q}^*}^{1} v_A(q) \frac{g(q)}{1 - G(\widehat{q}^*)} dq \equiv E[v_A(q) \mid q \geq \widehat{q}^*], \qquad (6) \\
P_B(\widehat{q}^*) &= \int_{0}^{\widehat{q}^*} v_B(q) \frac{g(q)}{G(\widehat{q}^*)} dq \equiv E[v_B(q) \mid q < \widehat{q}^*].
\end{aligned}
$$

Setting a price $p_n < P_n(\widehat{q}^*)$ is clearly suboptimal for the corresponding firm. This observation uses the restriction to passive beliefs, because changes in prices do not affect the customer's belief about advice, as captured by $\widehat{q}^*$; see, however, the discussion in online Appendix 1. Also, firms cannot profitably deviate by setting a sufficiently low price that induces the customer to always buy, irrespective of the adviser's recommendation. The information conveyed through advice is necessary for trade by our assumption (3) that the customer's *unconditional* expected valuation is below product cost. We comment on this restriction after characterizing the equilibrium.

Commissions. At $t = 1$, when commissions are set, firms' expected profits are given by

$$
\begin{aligned}
\pi_A &= [p_A - f_A - c_A][1 - G(q^*)], \qquad (7) \\
\pi_B &= [p_B - f_B - c_B] G(q^*).
\end{aligned}
$$

Profits depend directly on the actual cutoff q^*, which from (5) is a function of the actual commissions f_n chosen by firms. Given that the customer decides on the basis of the adviser's recommendation, the marginal customer type q^* is determined by the adviser's indifference condition (5). In addition, given that firms optimally set $p_n = P_n(\widehat{q}^*)$ according

to (6), profits depend on the expected cutoff $\widehat{q}^*$ and, thus, on the expected commissions that are anticipated by the customer.

Differentiating firms' profits, we obtain firms' best responses for given $\widehat{q}^*$ as follows. For firm A, we have

$$f_A = p_A - c_A - 2w\frac{1 - G(q^*)}{g(q^*)}, \tag{8}$$

when this is both strictly positive (otherwise $f_A = 0$) and not above $f_B + w$ (otherwise $f_A = f_B + w$). For firm B, we have

$$f_B = p_B - c_B - 2w\frac{G(q^*)}{g(q^*)}, \tag{9}$$

when this is both strictly positive (otherwise $f_B = 0$) and not above $f_A + w$ (otherwise $f_B = f_A + w$). Both best responses are unique by the hazard rate conditions (1) and (2).

Intuitively, incentives to pay commissions are higher when the firm's margin is higher. In fact, according to firms' best responses, an increase in the price is reflected in a one-for-one increase in the respective commission. Next, recall that the responsiveness of the adviser's recommendation to changes in commissions is given by $|dq^*/df_A| = dq^*/df_B = 1/(2w)$. Because an increase in commissions must be paid not only when a sale is made for $q = q^*$, but also for all inframarginal sales, firms' marginal costs from raising commissions are given by $G(q^*)$ for firm B and by $1 - G(q^*)$ for firm A. This is reflected in the last term in (8) and (9), respectively. The firms' trade-off between pushing sales, thereby capturing the marginal type q^*, and reducing their margin is analogous to the classic trade-off between price and quantity in the theory of oligopoly. There, a lower price results in higher sales but reduces the margin on all sales, including the inframarginal sales that would also have been made without a price cut. A symmetric trade-off holds for firm B, where inframarginal sales are given by $G(q^*)$.

The following immediate observations are also analogous to those made in oligopoly theory (cf. Helmut Bester, 1992). Best-response functions are strictly increasing so that commissions are strategic complements. This means that firm n finds it more profitable to raise its own commission f_n when it expects the rival firm n' to choose a higher commission $f_{n'}$. Finally, the hazard rate conditions (1) and (2) ensure that the two upward-sloping best responses intersect only once.

Equilibrium.   The preceding characterization of the firms' choice of commissions is valid for given customer expectations and product prices. In equilibrium, commissions as well as prices must be determined jointly; firms' choices of commissions must be optimal for

given customer expectations $\widehat{f}_n$, while customer expectations must be satisfied: $\widehat{f}_n = f_n$ for $n = A, B$ and, thus, $\widehat{q}^* = q^*$. Proposition 1 shows that these conditions jointly pin down a unique equilibrium outcome. We denote the equilibrium values of commissions, prices, and advice cutoff for the baseline case without disclosure by f_n^{ND}, p_n^{ND}, and q^{ND}.

When the adviser is less concerned about the suitability of the recommendation, firms' incentives to raise commissions are enhanced through two channels. First, the incentives to steer the adviser are increased because the advice becomes more responsive to commissions. Second, given the observation that firms' strategies to set commissions are strategic complements, the increase of one firm's commission has an additional feedback effect on the incentives of the other firm. Even though in equilibrium commissions thus change with w, in the symmetric case in which firms are equally cost-efficient ($c_A = c_B$) we always have $q^{ND} = 1/2$. When firms have the same incentives to steer advice, competition thus creates balanced incentives for the adviser, irrespective of the extent of the adviser's concern for suitability and, thus, irrespective of the level of commissions that prevails in equilibrium.

When instead firm A is more cost-efficient, its incentives to pay commissions and thereby expand sales are higher, so that $q^{ND} < 1/2$ results in equilibrium. Note that because an increase of a firm's commission must also be paid inframarginally (i.e., for all $q > q^{ND}$ for firm A and for all $q < q^{ND}$ for firm B), it becomes increasingly more costly for firm A to expand sales when its market share is already large. This dampening effect is more pronounced the higher is the adviser's concern for suitability, which is why an increase in w reduces the market share of the more cost-efficient firm.

To see this formally, when commissions are positive, we can substitute the respective first-order conditions for f_n^{ND} into the definition of the cutoff q^* in (5). After substituting for the equilibrium prices, we obtain that the equilibrium cutoff q^{ND} must satisfy

$$\left[E[v_A(q) \mid q \geq q^{ND}] - c_A\right] - \left[E[v_B(q) \mid q < q^{ND}] - c_B\right] = w\left[(1 - 2q^{ND}) + 2\frac{1 - 2G(q^{ND})}{g(q^{ND})}\right].$$
$$(10)$$

This equation determines how equilibrium market shares depend on supply-side differences. The left-hand side of (10) represents the difference in firms' margins (gross of commissions) and, thus, in their marginal benefits to steer advice through commissions. The right-hand side corresponds to the difference in marginal costs that affect the responsiveness of advice, and it comprises two terms. The first term, $w(1 - 2q^{ND})$, represents the adviser's preference to give suitable recommendations at the margin. The second term relates to the above-mentioned trade-off that firms face when marginally increasing commissions, given that the incremental commission must also be paid on inframarginal sales.

14

Proposition 1 *Once attention is restricted to pure-strategy equilibria with passive beliefs and to informative equilibria (when they exist), in the baseline scenario without disclosure there exists a unique equilibrium. When the adviser is less concerned about suitability (lower w), commissions of both firms increase, and strictly so when they are already positive. If firms are equally cost-efficient ($c_A = c_B$), the symmetric outcome $q^{ND} = 1/2$ always arises irrespective of w. If instead $c_A < c_B$, the market share of the more cost-efficient firm A increases (lower q^{ND}) when w decreases, and strictly so when commissions are positive.*

Firms' Profits. The proof of Proposition 1 also contains conditions for when commissions are strictly positive. Intuitively, this is the case when w is low so that the advice is sufficiently responsive to commissions. Then, substituting from firms' best responses, we obtain the equilibrium profits:

$$\pi_A^{ND} = 2w\frac{\left[1 - G(q^{ND})\right]^2}{g(q^{ND})}, \tag{11}$$
$$\pi_B^{ND} = 2w\frac{\left[G(q^{ND})\right]^2}{g(q^{ND})}.$$

In case of symmetry with $c_A = c_B$ and, thus, $q^{ND} = 1/2$, profits further simplify to $\pi_n^{ND} = w/[2g(1/2)]$. When advice is more responsive (lower w) or $q = q^{ND} = 1/2$ is more likely (higher $g(1/2)$), it becomes more attractive to increase commissions and so competition intensifies and profits decrease. Next, with asymmetric firms ($c_A < c_B$), from Proposition 1 we have $q^{ND} < 1/2$ and $dq^{ND}/dw > 0$, so that the market share of firm B increases when advice becomes less responsive. Combining this observation with the hazard rate assumption (2), we conclude that firm B's profits, as given in (11), are strictly increasing in w also when market shares are asymmetric. However, this need not be the case for the larger firm A, which suffers from a reduction in market share as advice becomes less responsive. The preceding observations again mirror those obtained in the standard Hotelling model in which firms compete in prices for final customers.

Recall that from (3) trade cannot occur without advice. We used this restriction to rule out the possibility that a firm can profitably deviate by sufficiently undercutting the rival and, thereby, induce the customer to buy its product even against the adviser's recommendation. We show in online Appendix 3 how condition (3) can be relaxed while still ensuring that such a deviation is not profitable. Alternatively, such a strategy would simply not be feasible for firms when the adviser could essentially prescribe a particular product, thereby leaving the customer only with the choice between buying this particular

product or no product at all.

If (3) does not hold and firms have direct access to customers, more generally an alternative to advised sales is to sell directly to the uninformed customers. Suppose that some firm n chooses such a regime of only direct sales, say at $t = 0$. If the adviser realizes the payoffs w_l and w_h only when the respective product is purchased through the intermediated channel, the adviser always wants to recommend the rival firm's product n'. Advice is then no longer informative, and even though firm n' may sell through the adviser, it optimally does not pay commissions. Both firms can then charge a price only equal to the average unconditional valuation $(v_l + v_h)/2$, given that firms' products are undifferentiated without the adviser's information. In the symmetric case with $c_A = c_B$, firms make zero profits, while with $c_A < c_B$ only the more cost-efficient firm A sells and obtains profits of $c_B - c_A$ in equilibrium, once we make the standard restriction that the less cost-efficient firm refrains from posting weakly dominated prices strictly below its cost c_B. While firm B would thus never choose direct sales, in the asymmetric case the choice for firm A is no longer immediate. In particular, the impact that the adviser's concern for suitability w has on firm A's profits is ambiguous. As we noted in Proposition 1, a lower value of w leads to more intense competition through commissions, but it generates a larger market share for firm A (cf. online Appendix 3 for further analysis).

4 Disclosure of Commissions

With disclosed commissions, the customer can directly infer the adviser's optimal choice of cutoff. We still look for an equilibrium in pure strategies in which advice is informative. Firm profits are now obtained from expressions (7), while noting that the customer's conditional expected valuations in (6) and, thus, the maximum prices now depend on the actual cutoff, q^*, rather than the expected cutoff, $\widehat{q}^*$. Hence, when an increase in the commission of one firm is observed by the customer, this firm is forced to reduce its price as the customer's conditional valuation decreases, while the rival can charge a higher price. Note that firms optimally set prices so as to extract the customer's full conditional valuations, $p_n = P_n(q^*)$, where the cutoff q^* now depends on the actual commissions.

Taking into account this dependence through differentiation of the respective condi-

tional valuations, we obtain the best responses with disclosure:

$$f_A = v_A(q^*) - c_A - 2w\frac{1 - G(q^*)}{g(q^*)}, \tag{12}$$

$$f_B = v_B(q^*) - c_B - 2w\frac{G(q^*)}{g(q^*)}.$$

Precisely, these conditions hold when they are nonzero and when $|f_A - f_B| \leq w$; otherwise, either $f_n = 0$ or $f_n = f_{n'} + w$, as explained in the proof of Proposition 2. As in the case without disclosure, we can show that best responses are unique, that they give rise to strategic complements, and that they intersect exactly once. We denote the respective equilibrium outcome with disclosure by f_n^D, p_n^D, and q^D.

Note also the direct analogy to the best responses without disclosure, as obtained in (8) and (9). The only difference is that prices p_n, which in equilibrium are equal to the respective conditional valuations, are now substituted by the corresponding marginal valuations, $v_A(q^*)$ and $v_B(q^*)$. Note that we obtain these conditions after substituting $p_n = P_n(q^*)$ into firms' profit functions and thereby taking into account how the customer's conditional valuations for the products change when advice is steered through commissions.

Given that the conditional valuations are strictly higher for both products than the marginal valuations whenever the adviser's cutoff is interior, we can already conclude that disclosure has a chilling effect on firms' incentives to pay commissions. In fact, the chilling effect of disclosure on each firm's incentives is further amplified in equilibrium by the fact that commissions are strategic complements.

When both commissions are positive, from substitution of the best responses into (5) we have that q^D solves

$$\left[v_A(q^D) - c_A\right] - \left[v_B(q^D) - c_B\right] = w\left[(1 - 2q^D) + 2\frac{1 - 2G(q^D)}{g(q^D)}\right]. \tag{13}$$

Again, this is uniquely determined and it is analogous to the characterization of q^{ND} in (10), once we replace the marginal valuations with the corresponding conditional valuations. We return below to a further comparison of the characterizations in (10) and (13).

Proposition 2 *Once attention is restricted to pure-strategy equilibria and to informative equilibria (when they exist), in the scenario with disclosure there exists a unique equilibrium. When the adviser is less concerned about suitability (lower w), commissions of both firms increase, and strictly so when they are already positive. If firms are equally*

cost-efficient ($c_A = c_B$), the symmetric outcome $q^D = 1/2$ always arises irrespective of w. If instead $c_A < c_B$, the market share of the more cost-efficient firm A increases (lower q^D) when w decreases, and strictly so when commissions are positive. Commissions are lower with disclosure than without disclosure, and strictly so when they are strictly positive without disclosure.

For both Propositions 1 and 2, the conditions that determine when commissions are strictly positive or equal to zero are reported in the proof. Intuitively, commissions are strictly positive whenever w is low so that advice is sufficiently responsive. In what follows, for ease of exposition we focus on the interior case with strictly positive commissions so that equilibrium cutoffs q^{ND} and q^D are characterized by (10) and (13), respectively.

Disclosure and Advice. Having established that disclosure dampens commissions, we now turn to the effect of disclosure on the advice cutoff. It is key to realize that this effect depends on the *relative* impact of disclosure on the commissions of the two competing firms. In the symmetric case with $c_A = c_B$, it follows immediately from Propositions 1 and 2 that competition always creates balanced incentives for the adviser, $q^{ND} = q^D = 1/2$. Thus, in this case disclosure has an impact only on the size of commissions, but leaves advice unaffected.

The case with asymmetric costs is more interesting. From Propositions 1 and 2 the market share of the more cost-efficient firm A is larger than that of the less cost-efficient firm B both with and without disclosure. We now show that while disclosure dampens the incentives of both firms to pay commissions, this effect is *relatively* stronger for the more cost-efficient firm, so that disclosure reduces firm A's market share.

Comparing the characterizations of the equilibrium cutoffs in (10) and (13), each firm's incentives to raise commissions depend on the respective *marginal* valuations for the case of disclosure and on the respective *average conditional* valuations for the case without disclosure. The comparison depends on the following key inequality:

$$E[v_A(q) \mid q \geq q^*] - v_A(q^*) > E[v_B(q) \mid q < q^*] - v_B(q^*), \tag{14}$$

which we now establish to hold for any given $q^* < 1/2$. This inequality means that the conditional valuation net of the marginal valuation is *relatively* higher for product A than for product B, provided that $q^* < 1/2$. Note that at $q^* = 1/2$ condition (14) holds with equality by our assumption that the distribution $G(q)$ of the adviser's belief is symmetric

around $q = 1/2$. At $q^* = 0$, instead, this condition holds strictly, because then the left-hand side equals $(v_l + v_h)/2 - v_l$, while the right-hand side is equal to zero, given that at $q^* = 0$ the marginal and the average valuations for product B are obviously the same. As we verify in the proof of Proposition 3, our monotone hazard rate conditions on the belief distribution guarantees that $E[v_A(q) \mid q \geq q^*] - v_A(q^*)$ is everywhere strictly decreasing and that $E[v_B(q) \mid q < q^*] - v_B(q^*)$ is everywhere strictly increasing in q^*, from which (14) follows when $q^* \neq 1/2$.

Proposition 3 *Disclosure of commissions does not affect advice when firms are equally cost-efficient, because then $q^D = q^{ND} = 1/2$. If instead firms' costs are asymmetric ($c_A < c_B$), the adviser recommends the less cost-efficient product B more often with disclosure than without, so that disclosure reduces the market share of the more cost-efficient firm A.*

5   Welfare and Policy

Turning to welfare analysis, in this section we specify that the adviser's concern for suitability arises only from an expected penalty (equal to w) following the sale of an unsuitable product.[10] When such penalties or fines are transfers, expected social welfare is maximized when the cutoff q^* is chosen to maximize

$$\omega = \int_0^{q^*} [v_B(q) - c_B] \, dG(q) + \int_{q^*}^1 [v_A(q) - c_A] \, dG(q).$$

Note that ω is strictly quasiconcave in q^* and maximized at $q^* = q_{FB}$ with

$$0 < q_{FB} := \frac{1}{2} - \frac{c_B - c_A}{2 (v_h - v_l)} < 1. \tag{15}$$

Intuitively, the first-best cutoff is $1/2$ when firms are equally cost-efficient, while otherwise $q_{FB} < 1/2$ is strictly decreasing when the cost difference $c_B - c_A > 0$ increases.

Proposition 4 *Suppose that the adviser's concern for suitability arises from a penalty w levied following the sale of an ex post unsuitable product. Then advice is always socially efficient when firms are symmetric, $c_A = c_B$, in which case $q^{ND} = q^D = q_{FB} = 1/2$. When*

---

[10]For example, violation of suitability regulation (such as FINRA/NASD Conduct Rule 2310 requiring brokers-dealers in the US to give suitable advice) results in fines and other disciplinary proceedings (such as expulsion), but not in compensatory damages to customers. Customers can typically obtain redress only if they are able to demonstrate fraud or breach of fiduciary duty, both of which typically entail much more stringent burdens of proof compared to suitability violations. See, for example, Norman S. Poser and James A. Fanto (2010).

firms are asymmetric ($c_A < c_B$) and the suitability concern is strictly positive ($w > 0$), with disclosure the market share of the more cost-efficient firm A is always too low: $q^D > q_{FB}$. Without disclosure, there exists a critical level of the penalty, w_{FB}, such that the market share of firm A is too low ($q^{ND} > q_{FB}$) when $w > w_{FB}$, is too high ($q^{ND} < q_{FB}$) when $w < w_{FB}$, and is efficient ($q^{ND} = q_{FB}$) when $w = w_{FB}$.

The outcome with symmetric firms follows immediately from our preceding observations. Regardless of whether commissions are disclosed and irrespective of the level of the penalty, symmetric competition to steer the adviser always leads to efficient advice.

Consider the asymmetric case with $c_A < c_B$. With disclosure, firms internalize how their commissions affect the adviser's recommendation and, thereby, the customer's conditional valuations for the products. Differences in firms' cost efficiencies result in differences in market shares to the extent that they are reflected in differences in commissions. From (13), note that with disclosure there are two reasons why differences in commissions fall short of differences in costs, resulting in $q^D > q_{FB}$. The first reason stems from the adviser's concern for suitability, which is now induced by the penalty. The second reason is that the incentives to increase commissions for the more cost-efficient firm are dampened by its larger market share, given that any increment in the commission has to be paid also on a larger stock of inframarginal sales compared to the less cost-efficient firm.

When w is sufficiently small, without disclosure the market share of the more cost-efficient firm is too large. In fact, from the characterization of q^{ND} in (10), we have that $q^{ND} \to q_L$ when $w \to 0$, where the limit is defined by

$$E[v_B(q) \mid q < q_L] - E[v_A(q) \mid q \geq q_L] = c_B - c_A.$$

The intuition for why the sales of product A "overshoot" is the following. Without disclosure there is no feedback mechanism that operates through the automatic reduction in the price extracted from the customer following an increase in commission. While the larger market share of the more cost-efficient firm still dampens its incentives to increase the commission, similar to the case with disclosure, without disclosure there is also a countervailing force. In order to maximize social welfare, firms' incentives to pay commissions should depend on the customers' marginal valuation at the resulting advice cutoff. In equilibrium without disclosure, instead, these incentives depend on the conditional average valuation. This difference between the conditional average valuation and the marginal valuation is strictly larger for firm A than for firm B, as we have shown in (14).

Corollary 1 *If firm A is more cost-efficient than firm B ($c_A < c_B$), there exists a threshold level $0 < w_D < w_{FB}$ for the penalty borne by the adviser when the customer purchases an ex-post unsuitable product, such that if the penalty is above that level, $w > w_D$, disclosure of commissions reduces efficiency; instead, if $w < w_D$, disclosure increases efficiency, while it does not affect efficiency if $w = w_D$.*

Optimal Penalty. When the adviser's concern for suitability is itself subject to policy, which penalty maximizes efficiency for a *given* disclosure regime? As we have seen, disclosure only affects efficiency when firms are differently cost-efficient. When commissions are not disclosed, from Proposition 4 we know that an increase in the adviser's penalty increases the efficiency of advice as long as $w < w_{FB}$, but decreases efficiency when already $w \geq w_{FB}$.

Instead, when commissions are disclosed, efficiency is increased by reducing the penalty w. In fact, substituting $w = 0$ into expression (13) results in $q^D = q_{FB}$, the efficient outcome. Through the price mechanism, mandatory disclosure induces firms to internalize the effect that steering the adviser's recommendation has on the quality of advice. In the presence of full disclosure, the imposition of penalties on the adviser is then counterproductive.

This stark implication, however, depends on a number of key assumptions. First, this result crucially depends on the ability of the customer to rationally infer the quality of advice from observed differences in commissions. Second, the result only holds when mandatory disclosure provides a perfect commitment for firms to reveal fully all incentives given to the adviser; however, policy makers often do not have complete control over disclosure.[11] Third, even though regulations that impose penalties and mandate disclosure should be *jointly* optimized, in many markets policy makers do not have complete control on penalties either, for example because these penalties and the adviser's concern for suitability are endogenous, as in Sections 6 and 7.

For the present analysis we have assumed that the penalty is not so large that the adviser prefers that the customer would sometimes *not* purchase. As noted above, by specifying that $w_l > w_0$ we ensure that there is effective competition by firms, because a higher commission then pushes up a firm's sales at the expense of the sales of the rival.

---

[11]Full disclosure of commissions is often difficult to implement in practice because firms have many opportunities to influence the adviser's recommendation through "soft" commissions, in the form of compensation for research, reimbursement of expense claims, and training courses in expensive resorts. See, for example, Financial Services Authority (2005).

In the more general case in which the adviser, instead, sometimes recommends to the customer not to purchase, firms behave like local monopolists. As we show in online Appendix 4, our main qualitative insights remain valid.

Caps on Commissions. Caps on commissions and other incentives are commonly imposed on various professional providers of health care and financial services (see footnotes 5 and 6). In our model, such a policy may mandate that $f_n \leq \overline{f}$, where $\overline{f}$ represents the cap imposed on commissions.

When firms are symmetric ($c_A = c_B$), any cap on commissions has no impact on the quality of advice, simply because the cap applies symmetrically to both firms and the firms have equally strong incentives to steer advice. Instead, when firms are asymmetric ($c_A < c_B$), advice can be affected by the imposition of a cap on commissions. A very low cap binds for both firms and, thus, constrains the outcome to be symmetric, with $q^{ND} = 1/2$ or $q^D = 1/2$, respectively. An intermediate cap binds only for the more cost-efficient firm A, which has higher incentives to pay commissions—thus the cap pushes the respective cutoffs q^{ND} or q^D closer to $1/2$ also in this case. On the other hand, the cap has no effect when it is so high that it is not binding for either firm.

Note the difference with a policy of increasing the penalty, which always reduces the incentives to pay commissions for *both* firms. Nevertheless, our qualitative insights from the analysis about the penalty also hold for caps on commissions. A binding cap always reduces efficiency when commissions are disclosed. Instead, when commissions are not disclosed, it is immediate to see that there exists $\overline{f}_{FB} > 0$ such that efficiency is lower both when the cap is decreased and when it is increased. We omit the derivations because they are analogous to those in Corollary 1.

6   Franchise Value and Reputation

Rather than imposing fines, a regulator or supervisor could also reduce the value of the adviser's franchise by revoking the business or professional licence. To analyze this possibility we now embed our model in a streamlined dynamic setting. At the beginning of each period, $\tau = 0, 1, ...$, a different customer arrives and demands a product. If the customer purchases the product, the realization of v is publicly observed at the end of the period. The adviser discounts profits by the factor $\delta < 1$. When $v = v_l$, we posit that the adviser loses the licence and is then replaced by another adviser, who then has the same access to customers. This setting ensures that revoking a licence has no effect

on efficiency. The benchmark of efficient advice is then still given by the cutoff q_{FB} from (15). To also abstract from additional complications that would arise if firms interacted with the adviser for more than one sale (cf., however, online Appendix 2), we stipulate that each period customers have different needs and that their choice is thus between two products of always different firms, even though each period is treated symmetrically.

Given stationarity of the problem, at the beginning of each period the adviser's expected utility is given by

$$u = \int_0^{q^*} [f_B + (1-q)U]\, dG(q) + \int_{q^*}^1 [f_A + qU]\, dG(q), \text{ where } U = \delta u. \qquad (16)$$

As can be seen immediately, the continuation value U now plays the role of w in our preceding analysis, including in the definition of the cutoff q^*. Note that we now stipulate that there is no additional liability. This allows us to apply the characterization results from Sections 3 and 4, simply by substituting $w = U$. Differently from the case with an exogenous penalty, when we compare the cases with and without disclosure, we also have to adjust $w = U$, which is now endogenous and depends on the expected future commissions that the adviser earns. Denote the respective values without and with disclosure by U^{ND} and U^D. Interestingly, when expected *future* commissions are higher, the adviser's recommendation becomes now less sensitive to commissions because more is at stake when the adviser loses the licence and, thus, the franchise value.

Proposition 5 *Suppose that the adviser's concern for suitability arises endogenously from the risk of losing the franchise following the realization of $v = v_l$, given that the licence is then revoked. When future profits become relatively unimportant (lower δ), commissions increase. When firms' costs are asymmetric ($c_A < c_B$), there exists a threshold level $\delta_D > 0$, such that disclosure becomes more efficient than non-disclosure when $\delta < \delta_D$, while disclosure reduces efficiency when $\delta > \delta_D$.*

When δ is low and so the adviser is not very motivated to give suitable advice because future profits and, thus, the value of the adviser's franchise are low, the market share of the more cost-efficient product A is always too large. Then, it is efficient to mandate disclosure to dampen firm A's relatively higher incentives to steer advice through commissions. This result is analogous to our earlier observations in Proposition 4 that disclosure of commissions improves efficiency when the penalty w is small. Instead, when δ is high and the adviser thus cares more about future business, we can show that both with and

without disclosure the threat of losing the franchise is always too high—$q^D > q_{FB}$ *and* $q^{ND} > q_{FB}$.

When determining which advice to give, the adviser trades off higher commissions in the present period with the prospect of losing future profits if the advice turns out to be unsuitable. Present profits may also weigh in more relative to future profits (low δ) when the market is expected to shrink in the future, when competition is anticipated to intensify, or during the transition phase in which a previously nationalized market is opened to competition.[12]

While so far we have specified that the adviser loses the franchise following the revocation of the licence by a regulator, our analysis also applies if the penalty follows from a loss in market reputation. For example, we can stipulate that only positive feedback (corresponding to the realization of v_h) keeps the adviser in the market. Unless v_h is observed, customers expect to receive a perfectly uninformative recommendation from the adviser in the future, and the adviser no longer expects to make any profits from commissions in the market. In this way, we can also guarantee that the adviser always prefers to recommend a purchase rather than no purchase at all.

7   Professional Concern

Even when we abstract from liability or the threat of losing future business, concern for suitability arises when the adviser personally cares about the well-being of the customer. This case may apply to members of particular professions (such as physicians) who care about the customer's well-being (whether or not the patient recovers), as captured by the suitability of the consumed product. Precisely, by assigning some weight γ on the respective realizations, we have $w_l = \gamma v_l$, $w_h = \gamma v_h$, and $w = \gamma(v_h - v_l)$. When we take this into account for welfare, efficiency is now maximized when the adviser's cutoff is equal to

$$\widetilde{q}_{FB} = \frac{1}{2} - \frac{c_B - c_A}{2(1 + \gamma)(v_h - v_l)}. \tag{17}$$

In contrast to the previous definition of q_{FB} in (15), with this professional specification the adviser's suitability concern enters welfare and, thus, the first-best cutoff.

---

[12]A case in point is the liberalization of the UK pension market at the end of the eighties which led to a ramp-up of commissions and ensuing allegations of unsuitable advice and egregious misselling in the nineties. As also stressed by Black (1997) and Black and Nobles (1998), among the many other factors at work, a key role is played by the development and implementation of effective suitability rules by regulators.

Consider first the disclosure regime. We observed previously that with asymmetric costs $(c_A < c_B)$ there are two forces (corresponding to the two terms on the right-hand side of expression (13)) that push the prevailing cutoff q^D strictly above the previous first-best cutoff q_{FB}. In our previous formulation, the agent's concern for suitability w was simply a transfer, which created a wedge between the equilibrium outcome and the efficient outcome. When, instead, w is part of the welfare criterion and $\widetilde{q}_{FB}$ in (17) results, this wedge is no longer present. However, we still have $q^D > \widetilde{q}_{FB}$ because of the agency problem between firms and the adviser, by which the more cost-efficient firm A, which has a larger market share, has higher (inframarginal) costs of marginally steering the agent.

Proposition 6 *Suppose that the adviser's concern for suitability arises from placing weight γ on the suitability of the customer's choice. Disclosure of commissions increases efficiency when γ is low, but decreases efficiency when γ is high.*

Proposition 6 shows that the results from Corollary 1 and Proposition 5 are robust also when the adviser's concern for suitability arises only from professional standards. However, while we obtain in Corollary 1 and Proposition 5 a unique cutoff on the penalty w or the importance of future profits δ, so that either regime is more efficient above or below this cutoff, Proposition 6 makes only an unambiguous comparison for sufficiently high or low values of γ.

8 Information Quality

The quality of the purchase decision made by the customer depends ultimately not only on the advice cutoff, but also on the quality of the adviser's information. To endogenize the adviser's information, consider a family of distribution functions $G(q; a)$, where $a \in [\underline{a}, \overline{a}]$ is real valued. It is convenient to suppose that G is everywhere continuously differentiable in q and a and that it always has full support $q \in [0, 1]$. We stipulate that a higher value of a rotates the distribution of the posterior belief q around $G(1/2; a) = 1/2$ through a mean-preserving spread with

$$\frac{dG(q; a)}{da} \gtreqless 0 \text{ for } q \lesseqgtr \frac{1}{2} \tag{18}$$

over $q \in (0, 1)$. A signal structure that results in such a rotation in the posterior distribution is more informative in the sense of Blackwell.[13] It proves convenient to work

---

[13]For more on rotations, see Johnson and Myatt (2006) and Dezsö Szalay (2009); and for the relation between integral precision and Blackwell sufficiency for dichotomies, see Ganuza and Penalva's (2010) Theorem 2.

directly with the distribution of posterior beliefs, given that it is well known that posterior beliefs can be equivalently derived from private signals. Note also that, as an immediate implication of the mean-preserving spread, the density at the mean, $g(1/2; a)$, is strictly decreasing in a.

Given our focus on the quality of the adviser's information, rather than on the adviser's potential bias as expressed by $q^* \neq 1/2$, we consider only the case with equally cost-efficient firms ($c_A = c_B$), which always results in a symmetric equilibrium cutoff. We further suppose that the hazard rate condition (1) is still satisfied for all distributions in the family $G(q; a)$. Online Appendix 5 reports a flexible analytical example in which we show how the hazard rate condition (1) is always satisfied by adequately choosing the upper bound $\bar{a}$ on information quality.

We are now interested in the adviser's incentives to invest in training or qualification that allows the provision of better-quality advice. For this purpose we stipulate that a is observably chosen at the beginning before contracting with firms, say at some time $t = 0$. The choice of a entails a private cost $k(a)$, where $k(a)$ is twice continuously differentiable with $k'(\underline{a}) = 0$ and $k(a) \to \infty$ as $a \to \bar{a}$. We discuss below the case in which information quality also depends on effort that is exerted only after the adviser is matched with a particular customer.

The adviser's incentives to become better qualified depend both on the concern for suitability and on how this affects the resulting monetary payoff taking into account the commissions paid by firms. To understand the impact of information quality on the equilibrium level of commissions, it is convenient first to analyze how information quality impacts firms' profits.

Information Quality and Firm Profits. When costs are symmetric, from expressions (11) for the case with undisclosed commissions we obtain that the equilibrium profits for each firm are

$$\pi^{ND} = \frac{w}{2g(1/2; a)}. \tag{19}$$

Similarly, by substituting for the first-order conditions we find that each firm's profits with disclosure are

$$
\begin{aligned}
\pi^D &= \pi^{ND} + \left[E[v_A(q) \mid q \geq q^D] - v_A(q^D) \right] \left[1 - G(q^D) \right]. \tag{20} \\
&= \pi^{ND} + (v_h - v_l) \int_{1/2}^1 \left[1 - G(q; a) \right] dq.
\end{aligned}
$$

Firm profits are strictly higher with disclosure. The difference is made up exactly by the difference between customers' conditional average valuation and their marginal valuation at the symmetric cutoff $1/2$. This is captured in the second term in expression (20). From our information quality condition (18), this term is indeed strictly larger when the adviser becomes better informed.

When commissions are not disclosed, firms fully compete away, through higher commissions, any increase in customers' valuation that arises when the adviser becomes better informed. Recall that without disclosure firm profits in (19) are only a function of the intensity with which they compete for the marginal type $q = 1/2$. The intensity of competition for the marginal type decreases when $g(1/2; a)$ is smaller, i.e., when it is less likely that the adviser has not updated the prior belief and is thus still fully uncertain about which product is more suitable. Hence, when commissions are not disclosed, firms benefit when the adviser becomes better informed, because the improvement in information reduces competition for the *marginal* customer. This effect is also present with disclosure, but firms then also benefit from the increase in the customer's *conditional* valuations when the advice is based on better information.

Adviser Incentives. The adviser's expected payoff when commissions are not disclosed is

$$u^{ND} = w_l + \int_0^{1/2} \left[f_B^{ND} - wq \right] g(q; a)dq + \int_{1/2}^1 \left[f_A^{ND} - w(1-q) \right] g(q; a)dq - k(a). \quad (21)$$

After integration by parts and substitution from the definition of the equilibrium cutoff $q^D = 1/2$ and $f_n^{ND} = f^{ND}$, this becomes

$$u^{ND} = (w_l + f^{ND}) + w \left[\int_0^{1/2} G(q; a)dq - \int_{1/2}^1 G(q; a)dq \right] - k(a). \quad (22)$$

We can obtain an analogous expression for the adviser's resulting payoff with disclosure

$$u^D = u^{ND} - (f^{ND} - f^D) = u^{ND} - 2(v_h - v_l) \int_{1/2}^1 \left[1 - G(q; a) \right] dq - k(a).$$

As is immediate, the difference in the adviser's payoff with and without disclosure exactly matches the difference in the sum of firms' profits between these two regimes (cf. expression (20)).

When the information quality a increases, there is clearly a direct positive effect on the payoff the adviser derives from the concern for suitability. This is captured by the

27

fact that the second term in (22) in brackets increases strictly by (18). In addition, a has an indirect effect on the adviser's payoff, through the induced change in the respective commissions f^{ND} and f^D. This effect is just the opposite of the effect on firms' profits, as discussed above. That is, with disclosure commissions depend only on the intensity of competition for the marginal customer $q = 1/2$, and they are thus higher when the adviser is less informed so that $g(1/2; a)$ is larger. Without disclosure, instead, there is a countervailing effect, because an increase in the customer's conditional valuations results in an increase in product prices, which is then passed through one-for-one into higher commissions. When commissions are not disclosed, the adviser fully extracts the benefits that better information quality has for customers.

Taking Section 5's efficiency criterion for the case in which the adviser's concern for suitability arises from regulatory penalties, we have:

Proposition 7 *Suppose that the adviser can observably invest at $t = 0$ in information quality a in the case with symmetric costs ($c_A = c_B$). The quality of information is strictly higher when commissions are not disclosed. The resulting level remains below the level that maximizes efficiency, unless the penalty w is sufficiently high.*

When commissions are not disclosed, the adviser's payoff becomes more responsive to the quality of information, given that the higher prices that firms can charge when customers can expect to receive advice that is more suitable are then passed on into higher commissions. This is not the case when commissions are disclosed. Both with and without disclosure, however, there is a tendency to invest too little in information because a less-informed adviser invites more competition through commissions. At the marginal type $q = 1/2$, the benefits from expanding sales are larger for firms. To maximize efficiency, this would need to be compensated through a sufficiently high penalty w, because this directly increases the adviser's concern for suitability. When w is sufficiently large, however, the adviser's choice of a becomes excessively high.

Note finally that the effect that a change in the quality of the adviser's information has on equilibrium commissions would be absent if, instead, information quality were chosen only later in the game, once commissions have already been set. For concreteness, suppose that the adviser now chooses unobservable effort a only after matching with a customer. In this case, information acquisition is covert rather than overt, as in the analysis reported above. For this timing of moves, information quality no longer affects commissions; instead, now commissions could affect the adviser's choice of a. Note from expressions (21)

and (22) that commissions do not directly affect the impact that information quality has on the adviser's payoff, but only indirectly through their effect on the prevailing cutoff. When $q^D = q^{ND} = 1/2$ results regardless of disclosure (given $c_A = c_B$), we conclude that our result on information acquisition incentives in Proposition 7 continues to hold when the adviser could, in addition to investing in information quality in $t = 0$, also acquire additional, customer-specific information at a later stage. Online Appendix 6 provides a more detailed analysis.

9   Conclusion

This paper investigates how competition through commissions and hidden kickbacks affects the quality of advice received by customers and the resulting allocation of products. Our model predicts that disclosure leads to a reduction of commissions. This reduction is more pronounced for the firm that is more cost-efficient and thus has a higher market share. The impact of disclosure on welfare is ambiguous. Disclosure is welfare enhancing when the adviser is little concerned for suitability because in that case the cost-efficient firm's market share is too high without disclosure. When, instead, the adviser's suitability concern is sufficiently strong, disclosure reduces welfare. Furthermore, disclosure reduces the adviser's incentives to acquire information in advance.

Some recent experimental studies suggest that imposing mandatory disclosure of commissions may have additional drawbacks. James M. Lacko and Janis N. Pappalardo (2004) conjecture that disclosed commissions may prevent information-overloaded customers from adequately digesting other payoff-relevant facts. In another experimental study, Daylian M. Cain, George Loewenstein, and Don A. Moore (2005) argue that disclosure of bias may lead advisers to feel morally justified when deviating from professional standards, resulting in a reduction in the quality of advice. While such effects may be only transitory in nature, our analysis suggests that mandatory disclosure or other interventions to reduce commission levels may have ambiguous welfare implications even in the long term, after customers and advisers have adjusted their expectations through repeated experience.

In our model, a single adviser obtained information that was sufficient to judge the suitability of competing products. Alternatively, one could imagine that different advisers were (tied) experts for individual products, so that a customer must shop around to obtain a full picture. In this case, the recommendation received from other advisers would become private information for the customer, with repercussions also on the equilibrium communication strategy of the first adviser.

Products' characteristics could also be endogenized. Investments in cost reduction or quality improvement only pay off when firms have access to a sufficiently large fraction of the market. If policy measures make it more costly for firms to adequately incentivize agents, product innovation may be inefficiently hampered.

As in other contract-theoretic analyses, we expect our insights to apply also to unconditional gifts paid by product providers who interact repeatedly with the same adviser—but we leave a formal analysis of relational commissions to future research.[14] Commissions may also affect an agent's incentives to provide other services, such as acquiring new customers. In the common agency case in which one firm can free ride on the incentives that other firms provide to the agent to locate customers, we expect that the agent's effort to locate customers should be inefficiently low. Once again, the reduction in commissions brought about by disclosure could further worsen this inefficiency.

---

[14]From a legal standpoint, payments to doctors from pharmaceutical companies are treated suspiciously even when they are not explicitly contingent on sales. In a recent and widely publicized settlement, the US Department of Justice contended successfully that various remunerations paid by AstraZeneca to doctors (to serve as authors of articles about uses of Seroquel, to travel to resort locations to advise AstraZeneca about marketing messages, and to give promotional lectures to other health care professionals) were actually intended to induce the doctors to prescribe Seroquel in violation of the federal Anti-Kickback Statute, 42 U.S.C. § 1320a-7b(b).

Appendix: Proofs

Proof of Proposition 1. Using (1), the best responses in (8) and (9) give rise to strategic complements and intersect at most once. As a function of $\widehat{q}^*$, f_A is increasing and f_B is decreasing, so that from (5) we obtain a nonincreasing and continuous function $q^*(\widehat{q}^*)$. That $q^*(0) > 0$ and $q^*(1) < 1$ follows from condition (3). Uniqueness of an equilibrium in pure strategies where $0 < q^{ND} < 1$ is then established from the requirement that $\widehat{q}^* = q^*$, and existence from (4).

We next characterize necessary and sufficient conditions for when one or both commissions are positive, making use of strict quasiconcavity of firms' programs for given expectations and thus prices. First, consider the commission paid by firm B. Differentiating firm B's profits with respect to f_B and evaluating at $f_B = 0$ gives

$$\left. \frac{d\pi_B}{df_B} \right|_{f_B=0} = [P_B(\widehat{q}^*) - c_B] \frac{g(q^*)}{2w} - G(q^*). \tag{23}$$

By quasiconcavity, a sufficient condition for $f_B^{ND} > 0$ to hold requires that (23) be strictly positive. Using monotonicity of $P_B(\hat{q}^*)$, $G(1/2) = 1/2$, and condition (2) gives

$$w < \tilde{w} \equiv g(1/2) \left[E[v_B(q) \mid q < 1/2] - c_B \right]. \tag{24}$$

Similarly, differentiating firm A's profits with respect to f_A and evaluating at $f_A = 0$ gives

$$\left. \frac{d\pi_A}{df_A} \right|_{f_A=0} = [P_A(\widehat{q}^*) - c_A] \frac{g(q^*)}{2w} - [1 - G(q^*)]. \tag{25}$$

Using quasiconcavity again, we obtain that $f_A^{ND} = 0$ if and only if $d\pi_A/df_A|_{f_A=0} \leq 0$, which occurs when $w \geq w' = g(1/2) \left[E[v_A(q) \mid q > 1/2] - c_A \right]$. Since $c_A \leq c_B$ and $E[v_A(q) \mid q > 1/2] = E[v_B(q) \mid q < 1/2]$ by (4), we have $w' \geq \tilde{w}$. In this case, q^{ND} is characterized by (10).

Finally, a threshold for w that provides also a necessary condition for when $f_B^{ND} > 0$ holds is obtained as follows. Substituting $f_B = 0$ into (8), we obtain from the resulting best response f_A a value for the cutoff q^*. The condition when $f_B^{ND} > 0$ is then obtained from evaluating the best response f_B in (9) at the resulting cutoff $\widehat{q}^* = q^*$. Given that both best responses are decreasing in w and that strategic complementarity holds, we thereby obtain a second cutoff $w'' < w'$, as long as $c_B > c_A$. When $c_A = c_B$, we have $w' = w''$, which is also obtained when (24) is satisfied with equality, noting that $E[v_A(q) \mid q > 1/2] = E[v_B(q) \mid q < 1/2]$.

The comparative statics of q^{ND} in w follows immediately from monotonicity of the hazard rate (1) and the reverse hazard rate (2).

Finally, the comparative statics of f_n^{ND} in w results from the following argument. Given that $dq^{ND}/dw > 0$, implying also that $P_B(q^{ND})$ decreases in w, from (9) and the hazard rate assumption we conclude that f_B^{ND} decreases. So as to still ensure that q^{ND} increases, also f_A^{ND} must decrease in w.

Proof of Proposition 2. Uniqueness of best responses in (12) and the fact that they intersect at most once both follows again from monotonicity of the hazard rate and the reverse hazard rate, as well as monotonicity of $v_A(q)$ and $v_B(q)$. Strategic complementarity of commissions follows from

$$\frac{df_A}{df_B} = \frac{v_A'(q^*) - 2w \frac{d}{dq^*} \frac{1-G(q^*)}{g(q^*)}}{2w + v_A'(q^*) - 2w \frac{d}{dq^*} \frac{1-G(q^*)}{g(q^*)}} > 0; \quad \frac{df_B}{df_A} = \frac{v_B'(q^*) - 2w \frac{d}{dq^*} \frac{G(q^*)}{g(q^*)}}{-2w + v_B'(q^*) - 2w \frac{d}{dq^*} \frac{G(q^*)}{g(q^*)}} > 0.$$

Condition (3) again guarantees that $0 < q^D < 1$. When $f_n^D > 0$ for $n = A, B$, q^D is uniquely pinned down by (13), satisfying $q^D = 1/2$ when $c_A = c_B$, and $q^D < 1/2$ along with $dq^D/dw > 0$ when $c_A < c_B$. A sufficient condition for $f_n^D > 0$ is that

$$w < g(1/2)\left[v_B(1/2) - c_B\right].$$

In analogy to the analysis in the proof of Proposition 1, to obtain necessary and sufficient conditions for when either commission is strictly positive, we can derive two cutoffs $0 < w'' \leq w'$. The fact that $f_A^D > 0$ holds if and only if $w < w' = g(1/2)\left[v_A(1/2) - c_A\right]$ follows from strict quasiconcavity of π_A and the respective best response. The cutoff w'', with $w'' < w'$ when $c_A < c_B$, is obtained from the best response of B after substituting for q^* the value derived with $f_B = 0$ and with f_A given by the respective best response for A.

Next, the comparative statics of q^D and f_n^D follow from the same arguments as those for q^{ND} and f_n^{ND} in the proof of Proposition 1. It remains to show that commissions are strictly higher without disclosure when they are strictly positive. We show first that this must hold for at least one firm. To see this, suppose first that $q^{ND} \leq q^D$. Comparing the respective best responses (9) and (12) for firm B, and also using (1) and $E[v_B(q) \mid q < q^{ND}] > v_B(q^D)$, we have that $f_B^{ND} > f_B^D$. When $q^{ND} > q^D$, instead, a symmetric argument implies $f_A^{ND} > f_A^D$. Next, we show that indeed both $f_n^{ND} > f_n^D$ when strictly positive. To see this, suppose $f_A^{ND} > f_A^D$, as we already established that this must hold for at least one firm. Thus, arguing to a contradiction, $f_B^{ND} \leq f_B^D$ would then imply $q^{ND} < q^D$, but this would from comparing (9) with (12) imply also the opposite, namely $f_B^{ND} > f_B^D$. In case $f_B^{ND} > f_B^D$, we can contradict $f_A^{ND} \leq f_A^D$ by an analogous argument.

Proof of Proposition 3. It remains to establish condition (14) when $q^* \neq 1/2$. As noted in the main text, we do this by showing that $E[v_A(q) \mid q \geq q^*] - v_A(q^*)$ is everywhere strictly decreasing, while $E[v_B(q) \mid q < q^*] - v_B(q^*)$ is everywhere strictly increasing. By symmetry of $G(q)$, it is enough to establish the property for product B, for which integration by parts yields

$$E[v_B(q) \mid q < q^*] - v_B(q^*) = (v_h - v_l) \int_0^{q^*} \frac{G(q)}{G(q^*)} dq. \tag{26}$$

Using $H(q^*) = \int_0^{q^*} G(q) dq$ and thus $H'(q^*) = G(q^*)$, showing that expression (26) is strictly increasing is thus equivalent to showing that $H'(q)/H(q)$ is strictly decreasing for all q, i.e., that $H(q)$ is logconcave. By Mark Bagnoli and Ted Bergstrom's (2005) Theorem 1, this property is implied by logconcavity of $G(q^*)$, which in turn is equivalent to the decreasing reverse hazard rate condition (2), which immediately follows from (1) and symmetry of $G(q)$.

Proof of Proposition 4. The case with $c_A = c_B$ follows from the discussion in the main text. Suppose thus that $c_A < c_B$. That $q^D > q_{FB}$ holds with disclosure follows from condition (13), which pins down q^D. Precisely, recall that the left-hand side of (13) increases in q^D while the right-hand side decreases in q^D, and note that when substituting q_{FB} from (15) for q^D, the left-hand side is strictly lower than the right-hand side.

Without disclosure, recall first that q^{ND} is continuous and strictly increasing in w. Existence of w_{FB} follows then from the observation that $q^{ND} > q_{FB}$ surely holds for high w, while when $w \to 0$ we have from (10) that $q^{ND} \to q_L$ satisfying $E[v_B(q) \mid q < q_L] - E[v_A(q) \mid q \geq q_L] = c_B - c_A$. Finally, $q_L < q_{FB}$ follows from condition (14).

Proof of Corollary 1. Denote the welfare levels achieved at the equilibrium cutoffs without and with disclosure by ω^{ND} and ω^D, and welfare at the first-best cutoff by ω_{FB}. By monotonicity of $q^D > q_{FB}$, we have $d\omega^D/dw \leq 0$ (and strictly so when $q^D < 1/2$), while from inspection of (13) we have $q^D \to q_{FB}$ and thus $\omega^D \to \omega_{FB}$ when $w \to 0$. Next, recall from the proof of Proposition 4 that without disclosure $\omega^{ND} = \omega_{FB}$ holds when $w = w_{FB}$, while $d\omega^{ND}/dw > 0$ when $w < w_{FB}$, $d\omega^{ND}/dw < 0$ when $w > w_{FB}$, and $q^{ND} \to q_L < q_{FB}$ as $w \to 0$. Thus we obtain the asserted cutoff $w_D < w_{FB}$ at which $\omega^{ND} = \omega^D$, with $\omega^{ND} < \omega^D$ when $w < w_D$ and $\omega^{ND} > \omega^D$ when $w > w_D$.

Proof of Proposition 5. From maximization of (16) we obtain that, each period, the adviser optimally chooses q^*, as characterized by (5) with $w = U$. Each period, the two

firms maximize π_A and π_B. From these observations we can thus apply the characterization results in Propositions 1 and 2. Recall further that, given that there is no deadweight loss from replacing the adviser, the efficient cutoff is still given by q_{FB}.

Consider now the case with disclosure and $c_A < c_B$, where $q^D > q_{FB}$ follows as $w = U > 0$. That $dq^D/d\delta > 0$ follows next if and only if $dU/d\delta > 0$. We argue to a contradiction and suppose, instead, that U decreases when δ increases. But as $w = U$, we know from Proposition 2 that then f_A^{ND} and f_B^{ND} must both increase, which from (16), given that q^{ND} is chosen optimally by the adviser, must necessarily increase u and thus U—a contradiction. The same argument applies without disclosure, so that also $dq^{ND}/d\delta > 0$. Existence of a threshold $\delta_D > 0$ for δ so that $q^{ND} < q_{FB}$ for higher δ and $q^{ND} > q_{FB}$ for lower δ follows then from Corollary 1 and observing that $U = w \to 0$ as $\delta \to 0$.

Proof of Proposition 6. Given that an increase in γ is equivalent to an increase in w, both q^{ND} and q^D strictly increase, provided $c_A < c_B$; otherwise, $q^{ND} = q^D = 1/2$ for all γ. This follows from Propositions 1 and 2. Also, we then have $q^D > q^{ND}$ and $q^D > \widetilde{q}_{FB}$, albeit now also $\widetilde{q}_{FB}$ strictly increases in w and thus in γ. It remains to determine when $q^{ND} > \widetilde{q}_{FB}$ and when the opposite holds.

From the definitions of $\widetilde{q}_{FB}$ in (17) and q^{ND} in (10) we have that $q^{ND} < \widetilde{q}_{FB}$ if and only if

$$[E[v_A(q) \mid q \geq \widetilde{q}_{FB}] - v_A(\widetilde{q}_{FB})] - [E[v_B(q) \mid q < \widetilde{q}_{FB} - v(\widetilde{q}_{FB})] > 2w\frac{1 - 2G(\widetilde{q}_{FB})}{g(\widetilde{q}_{FB})}. \quad (27)$$

When w is sufficiently small, then (27) holds. To see this, note that the right-hand side of (27) goes to zero, while the left-hand side is bounded away from zero given that $\widetilde{q}_{FB} \to q_{FB} < 1/2$ as $w \to 0$. When w is sufficiently large, the converse holds strictly, so that $q^{ND} > \widetilde{q}_{FB}$.

Proof of Proposition 7. We consider the adviser's choice of a at $t = 0$ under the two disclosure regimes. From $du^{ND}/da \geq du^D/da$, which holds strictly when commissions are strictly positive, monotonicity of the corresponding sets of maximizers follows from a standard monotone comparative statics argument.

Next, efficiency is highest when the respective choice of a maximizes

$$\omega - k(a) = \widetilde{\omega} = (v_l - c) + (v_h - v_l)\left[\int_0^{1/2} G(q; a)dq - \int_{1/2}^1 G(q; a)dq\right] - k(a),$$

where we use $c_n = c$. From comparison with u^{ND} we have $du^{ND}/da < d\widetilde{\omega}/da$ for all sufficiently low w. Note that then $df^{ND}/da < 0$ holds strictly. By monotone comparative

statics we conclude again that any choice of a that maximizes u^{ND} lies strictly below any efficient choice of a.

References

American Medical Association. 1948. "Editorial: Rebates, Kickbacks, Commissions and Medical Ethics." *Journal of the American Medical Association*, 136(3): 176–177.

Bagnoli, Mark, and Ted Bergstrom. 2005. "Log-Concave Probability and its Application." *Economic Theory*, 26(2): 445-469.

Bar-Isaac, Heski, Guillermo Caruana, and Vicente Cuñat. 2010. "Information Gathering and Marketing." *Journal of Economics and Management Strategy*, 19(2): 375–401.

Basel Committee on Banking Supervision. 2008. *Customer Suitability in the Retail Sale of Financial Products and Services.* Bank for International Settlements.

Bernheim, B. Douglas, and Michael D. Whinston. 1986. "Menu Auctions, Resource Allocation, and Economic Influence." *Quarterly Journal of Economics*, 101(1): 1–32.

Bester, Helmut. 1992. "Bertrand Equilibrium in a Differentiated Duopoly." *International Economic Review*, 33(2): 433–448.

Biglaiser, Gary. 1993. "Middlemen as Experts." *RAND Journal of Economics*, 24(2): 212–23.

Black, Julia. 1997. *Rules and Regulators.* Oxford: Clarendon Press.

Black, Julia, and Richard Nobles. 1998. "Personal Pensions Misselling: The Causes and Lessons of Regulatory Failure." *Modern Law Review*, 61(6): 789-820.

Bolton, Patrick, Xavier Freixas, and Joel Shapiro. 2007. "Conflicts of Interest, Information Provision, and Competition in Banking." *Journal of Financial Economics*, 85(2): 297–330.

Cain, Daylian M., George Loewenstein, and Don A. Moore. 2005. "The Dirt on Coming Clean: Perverse Effects of Disclosing Conflicts of Interest." *Journal of Legal Studies*, 34(1): 1–25.

Center for Responsible Lending. 2009. "Auto Dealers' Lending Abuses Cost Billions: Car Loans Are Too Important to Be Excused from CFPA," CRL Issue Brief.

Colwell, Peter F., and Charles M. Kahn. 2001. "The Economic Functions of Referrals and Referrals Fees." *Journal of Real Estate Finance and Economics*, 23(3): 267–296.

Consumer Federation of America. 2004. *The Hidden Markup of Auto Loans: Consumer Costs of Dealer Kickbacks and Inflated Finance Charges.*

Cummins, David, and Neal Doherty. 2006. "The Economics of Insurance Intermediaries." *Journal of Risk and Insurance*, 73(3): 359–396.

Crawford, Vincent P., and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica*, 50(6): 1431–1451.

Durbin, Erik, and Ganesh Iyer. 2009. "Corruptible Advice." *American Economic Journal: Microeconomics*, 1(2): 220–242.

Financial Services Authority. 2005. "Bundled Brokerage and Soft Commission Arrangements for Retail Investment Funds." Consultation Paper 2005/13.

Financial Services Authority. 2008. "Transparency, Disclosure and Conflicts of Interest in the Commercial Insurance Market." Discussion Paper 2008/2.

Ganuza, Juan-Josè, and Josè S. Penalva. 2010. "Signal Orderings Based on Dispersion and the Supply of Private Information in Auctions." *Econometrica*, 78(3): 1007–1030.

Hart, Oliver D., and Jean Tirole. 1990. "Vertical Integration and Market Foreclosure." *Brookings Papers on Economic Activity*, 205–76.

Hotelling, Harold. 1929. "Stability in Competition." *Economic Journal*, 39(153): 41–57.

Inderst, Roman, and Marco Ottaviani. 2009. "Misselling through Agents." *American Economic Review*, 99(3): 883–908.

Jackson, Howell E., and Laurie Burlingame. 2007. "Kickbacks and Compensation: The Case of Yield Spread Premiums." *Stanford Journal of Law, Business & Finance*, 12(2): 289–361.

Johnson, Justin P., and David P. Myatt. 2006. "On the Simple Economics of Advertising, Marketing, and Product Design." *American Economic Review*, 96(3): 756–84.

Keith, Ernst, Debbie Bocian, and Wei Li. 2008. "Steered Wrong: Brokers, Borrowers, and Subprime Loans." Center for Responsible Lending.

Lacko, James M., and Janis N. Pappalardo. 2007. "Improving Consumer Mortgage Disclosures: An Empirical Assessment of Current and Prototype Disclosure Forms." Federal Trade Commission Bureau of Economics Staff Report.

Lewis, Tracy R., and David E. M. Sappington. 1994. "Supplying Information to Facilitate Price Discrimination." *International Economic Review*, 35(2): 309–27.

Li, Ming, and Kristóf Madarász. 2008. "When Mandatory Disclosure Hurts: Expert Advice and Conflicting Interests." *Journal of Economic Theory*, 139(1): 47–74.

Li, Wei. 2010. "Peddling Influence through Intermediaries." *American Economic Review*, 100(3): 1136–62.

Lizzeri, Alessandro. 1999. "Information Revelation and Certification Intermediaries." *RAND Journal of Economics*, 30(2): 214–231.

McAfee, R. Preston, and Marius Schwartz. 1994. "Opportunism in Multilateral Vertical Contracting: Nondiscrimination, Exclusivity, and Uniformity." *American Economic Review*, 84(1): 210–230.

McGuire, Thomas G. 2000. "Physician Agency." Chapter 9 in *Handbook of Health Economics, Volume 1, Part 1*, ed. Anthony J. Culyer and Joseph P. Newhouse, 461–536. Amsterdam: North Holland.

Millenson, Michael L. 2003. "Getting Doctors to Say Yes to Drugs: The Cost and Quality Impact of Drug Company Marketing to Physicians." BlueCross BlueShield Association.

Morgan, John, and Phillip C. Stocken. 2003. "An Analysis of Stock Recommendations." *RAND Journal of Economics*, 34(1): 183–203.

Moscarini, Giuseppe, and Marco Ottaviani. 2001. "Price Competition for an Informed Buyer." *Journal of Economic Theory*, 101(2): 457–93.

Owen, Bruce M. 1977. "Kickbacks, Specialization, Price Fixing, and Efficiency in Residential Real Estate Markets." *Stanford Law Review*, 29(5): 931–967.

Pahl, Cynthia. 2007. "A Compilation of State Mortgage Broker Laws and Regulations, 1996–2006." Federal Reserve Bank of Minneapolis, Community Affairs Report No. 2007-2.

Pauly, Mark V. 1979. "The Ethics and Economics of Kickbacks and Fee Splitting." *Bell Journal of Economics*, 10(1): 344–352.

Poser, Norman S., and James A. Fanto. 2010. *Broker-Dealer Law and Regulation.* 4th ed. New York: Aspen Publishers.

Schwarcz, Daniel. 2007. "Beyond Disclosure: The Case for Banning Contingent Commissions." *Yale Law & Policy Review*, 25(2): 289–336.

Szalay, Dezsö. 2009. "Contracts with Endogenous Information." *Games and Economic Behavior*, 65(2): 586–625.

Roman Inderst and Marco Ottaviani

"Competition Through Commissions and Kickbacks"

Online Appendix

1. Beliefs

For the case without disclosure we have specified passive beliefs, as commonly assumed in the literature on vertical contracting. That literature has also analyzed the following alternative specification with wary beliefs. When observing a deviating contract offer from a common supplier, a downstream firm tries to infer how the supplier should have *optimally* (and secretly) adjusted contracts offered to other, competing downstream firms. In what follows, we show that when we apply to our model a restriction on beliefs that is in the same spirit, we can support the same unique equilibrium outcome as in Proposition 1.

Hence, a customer who observes a deviating price from some firm, say $\widehat{p}_A \neq p_A$, tries to infer how firm A, in anticipation of this deviation, should have optimally adjusted its commission. When firm A still expects the customer to follow the adviser's recommendation, this commission maximizes π_A and thus is given by (8), where profits are now evaluated at $\widehat{p}_A$. Profits would be zero regardless of firm A's choice of commission if, instead, the customer no longer followed the adviser's recommendation to buy good A.

Formally, for a given candidate equilibrium, we thus specify the following refinement on beliefs. When observing a deviating price $\widehat{p}_A \neq p_A$, beliefs assign all probability to commissions $\widehat{f}_A$ that maximize $[\widehat{p}_A - \widehat{f}_A - c_A][1 - G(\widehat{q}^*)]$, where $\widehat{q}^*$ is obtained from substituting $\widehat{f}_A$ and f_B into (5); when $\widehat{p}_B \neq p_B$, beliefs assign all probability to commissions $\widehat{f}_B$ that maximize $[\widehat{p}_B - \widehat{f}_B - c_A]G(\widehat{q}^*)$, where $\widehat{q}^*$ is now obtained from substituting $\widehat{f}_B$ and f_A into (5). When both $\widehat{p}_A \neq p_A$ and $\widehat{p}_B \neq p_B$, then this still applies for each $\widehat{f}_n$ individually. (Recall that without disclosure firms do not observe each others' choice of commissions.)

With this specification of beliefs, we now set up firms' optimization problems. For this note first that from (8) we can define for each p_A a unique value $\widehat{f}_A(p_A)$ and, together with the (equilibrium) $f_B = \widehat{f}_B$, a perceived cutoff $\widehat{q}^*(p_A)$, which is increasing in p_A. That is, beliefs assign probability one to a single commission, $\widehat{f}_A(p_A)$. Consequently, a customer who holds these beliefs will only follow the recommendation to buy A if, using (6), $p_A \leq P_A(\widehat{q}^*(p_A))$.[15] Note that, for given f_B, there is thus a unique price p_A for which this holds with equality. Given that this upper bound is independent of the actually

---

[15]Note that the customer can rightly believe that the adviser still applies a cutoff $\widehat{q}^*$ when facing commissions $\widehat{f}_A$ and f_B in case $p_A \leq P_A(\widehat{q}^*)$ and $p_B \leq P_B(\widehat{q}^*)$.

chosen commission, it is optimal for firm A to set p_A so that $p_A = P_A(\widehat{q}^*(p_A))$. This implies, however, that it is indeed optimal to choose f_A according to (8), and that consequently $q^* = \widehat{q}^*$. Given that the same logic applies to firm B, the equilibrium is pinned down by the same requirements as with passive beliefs (Proposition 1).

2. Agency

Our analysis, similar to standard contracting problems in the literature, focuses on a product sale to a single customer. We discuss now to what extent this matters. To do so, we consider the opposite extreme with a very large number of customers. Precisely, suppose that a unit mass of customers arrives simultaneously, after commissions and prices have been specified, as previously. The adviser's belief about each customer is an independent draw from $G(q)$. Generally, firms can now condition their commission payments on the mass (quantity) of sales, q_A and q_B, where $q_A + q_B = 1$. We denote such payments by $F_A(q_A)$ and $F_B(q_B)$. It is immediate to see that by optimality the agent once again chooses a cutoff rule q^* and realizes $q_A = 1 - q^*$ and $q_B = q^*$. With passive beliefs that do not depend on p_n, firms' profits as a function of q^* are given by

$$
\begin{aligned}
\pi_A &= (p_A - c_A)[1 - G(q^*)] - F_A(1 - q^*), \\
\pi_B &= (p_B - c_B)G(q^*) - F_B(q^*).
\end{aligned}
$$

For given q^*, the adviser's utility is

$$
u = w_l + w \left[\int_0^{q^*} (1 - q)dG(q) + \int_{q^*}^1 qdG(q) \right] + F_A(1 - q^*) + F_B(q^*).
$$

We now restrict consideration to continuously differentiable payments F_A and F_B, so that we obtain from optimality of the adviser

$$
\frac{du}{dq^*} = wg(q^*)(1 - 2q^*) + F_B'(q^*) - F_A'(1 - q^*) = 0.
$$

From this, we immediately obtain firms' marginal cost of implementing a different cutoff q^*. Substituting this into the firms' respective first-order condition, as obtained from π_A and π_B, and using $p_A = P_A(q^*)$ and $p_B = P_B(q^*)$, we obtain for the unique equilibrium cutoff without disclosure, which we now denote by $\bar{q}^{ND}$,

$$
\left[E[v_A(q) \mid q \geq \bar{q}^{ND}] - c_A\right] - \left[E[v_B(q) \mid q < \bar{q}^{ND}] - c_B\right] = w(1 - 2\bar{q}^{ND}). \tag{28}
$$

Even though with continuous differentiability of F_A and F_B we have pinned down a unique equilibrium cutoff and thus unique market shares, firms' profits are not uniquely

pinned down. These depend, instead, on the level of F_A and F_B at the cutoff $\overline{q}^{ND}$. In equilibrium, $F_B(\overline{q}^{ND})$ and $F_A(1 - \overline{q}^{ND})$ must be set so that the adviser has no incentive to deviate non-locally, by choosing any other cutoff $q^* \neq \overline{q}^{ND}$. As is well known in the literature on common agency (cf. Bernheim and Whinston 1986), there are multiple equilibria, given that the off-equilibrium values of F_n are not payoff relevant for the respective firm n, but determine the adviser's outside option and thus affect the value of $F_{n'}$ that is necessary at $\overline{q}^{ND}$ to keep the adviser from deviating.

We next compare (28) with (10), which we obtained in the text for the baseline case with a single customer. The only difference is that the term $2\frac{1-2G(q^{ND})}{g(q^{ND})}$ on the respective right-hand side is now absent. Recall that this term relates to the need for each firm to pay the incremental commission also inframarginally (i.e., for all realizations $q < q^{ND}$ for firm B and all realizations $q > q^{ND}$ for firm A) when steering advice to capture an additional customer. For $c_A < c_B$ we thus have that $\overline{q}^{ND} < q^{ND}$. However, the comparative statics result from Proposition 1 still holds.

Proceeding analogously for the case with disclosure, we obtain for the equilibrium cutoff $\overline{q}^D$ the requirement

$$\left[v_A(\overline{q}^D) - c_A\right] - \left[v_B(\overline{q}^D) - c_B\right] = w\left(1 - 2\overline{q}^D\right). \tag{29}$$

Comparing now (29) with (13), we have again $\overline{q}^D < q^D$ when $c_A < c_B$. The comparative statics result from Proposition 2 still holds, and the comparison of disclosure with non-disclosure in Proposition 3 also applies.

Finally, we discuss the robustness of our normative implications. Consider first the cases analyzed in Sections 5 and 6. There, from the definition of q_{FB} we still have that $\overline{q}^D > q_{FB}$ when $c_A < c_B$, and it is then immediate that all results of these two sections still apply, once the thresholds for w are appropriately modified. For Section 7's application, where the adviser's concern for suitability is part of the welfare criterion, we have instead that $\overline{q}^D = \widetilde{q}_{FB}$, regardless of the value of γ.

3. Deep Undercutting and Direct Sales

In this appendix, we first analyze how the results of Proposition 1 extend when assumption (3) is relaxed, so that there is scope for trade even without advice because the unconditional average valuation exceeds at least the cost of firm A. In this case the possibility arises for firms to *deviate* by undercutting the rival's price so as to sell to the customer against the adviser's recommendation. In a second step we then expand on the discussion reported at

the end of Section 3 by analyzing the scope for direct sales without the adviser's information in equilibrium.

Deep Undercutting. Recall that assumption (3) ruled out the possibility that either firm can cut its price sufficiently so as to induce customers to purchase its product even against the recommendation of the adviser. As we noted in the main text, assumption (3) would no longer be needed in case such a strategy was ruled out as the customer was simply prevented from purchasing a product against the adviser's recommendation. The customer has then only the option of either purchasing the prescribed product or not purchasing at all—as is the case of a physician's prescription of a pharmaceutical product or medical device. For the case in which this restriction of the customer's choice is not feasible, however, we analyze next how condition (3) can be relaxed.

Note first that when (3) is relaxed, to ensure that still $0 < q^{ND} < 1$ holds for all w, so that both products are sold in equilibrium, the cost difference must not be too large:

$$c_B - c_A < \frac{v_h - v_l}{2}.$$

Given passive beliefs, the discussed deep undercutting does not strictly pay for either firm when

$$
\begin{aligned}
\pi_A^{ND} &\geq \int_0^{q^{ND}} v_A(q) \frac{g(q)}{G(q^{ND})} dq - c_A, \\
\pi_B^{ND} &\geq \int_{q^{ND}}^1 v_B(q) \frac{g(q)}{1 - G(q^{ND})} dq - c_B.
\end{aligned}
\tag{30}
$$

Note that both conditional valuations at the right-hand side of expressions (30) are strictly smaller than the unconditional valuation $(v_l + v_h)/2$. They capture the respective valuations of a customer who decides to buy a product against the adviser's recommendation. Conditions (30) relax the restriction imposed in (3). When these conditions do not hold and when the adviser cannot prescribe the purchase of a particular product, as discussed above, then no equilibrium in pure strategies satisfies our imposed restrictions (of passive beliefs and of an informative equilibrium).

Direct Sales. As discussed in the main text, when either firm chooses in $t = 0$ to only directly sell its product, then advice is always uninformative in the equilibrium of the resulting subgame. When (3) holds, profits for both firms are zero. And when (3) does not hold, they are still always zero for firm B and again zero for both firms when $c_A = c_B$.

43

Hence, to see whether direct sales can be an outcome of the expanded game, we only have to consider profits for firm A in case $c_A < c_B$ and when (3) does not hold.

Given that direct sales lead to (undifferentiated) Bertrand competition, we have profits of $c_B - c_A$ for firm A. Recall that when (3) does not hold, we have $c_B \leq \frac{v_h + v_l}{2}$. Hence, the considered deviation of firm A to direct sales is strictly profitable when

$$\pi_A^{ND} \geq c_B - c_A, \tag{31}$$

given that profits are given by π_n^{ND} when both firms choose intermediated sales with advice, assuming that conditions (30) hold. Generally, conditions (30) do not imply (31) and vice versa. In particular, note that when, holding all else constant, we change only the less cost-efficient firm's cost c_B, condition (31) becomes stricter while conditions (30) are relaxed. In particular, when c_B becomes low enough to be equal to c_A, we know that (31) is always satisfied, even though (30) may not hold.

4. Not Fully Covered Market

In the main text, we restricted consideration to the case in which, both on and off the equilibrium path, the adviser's recommendation was essentially restricted to that of either product A or product B. This followed from our specification that $w_l > w_0$, and it ensured that firms always were in direct competition through commissions. An increase in one firm's commission, even if only marginal, pushes up the firm's market share at the expense of the rival. In the terminology of Hotelling competition, the market is fully covered. In what follows, we consider the case in which the market is not necessarily fully covered. To be specific, we suppose that the adviser is subject to a penalty following an unsuitable sale but obtains no direct benefits from a sale: $w_h = w_0 = 0$ and $w = -w_l > 0$.

When, for given commissions, the adviser prefers that the customer sometimes does not purchase any product, two cutoffs arise, $0 \leq q_B^* \leq q_A^* \leq 1$. The adviser recommends a purchase of B when $q \leq q_B^*$, a purchase of A when $q > q_A^*$, and no purchase when $q_B^* < q < q_A^*$. Provided that the customer follows the respective recommendations and provided that cutoffs are interior, we have

$$
\begin{aligned}
q_A^* &= \frac{(w_0 - w_l) - f_A}{w} = 1 - \frac{f_A}{w}, \\
q_B^* &= \frac{(w_h - w_0) + f_B}{w} = \frac{f_B}{w}.
\end{aligned}
$$

Note that each cutoff depends only on the commission of the corresponding firm.

Consider first the case with disclosure, where we obtain a unique pair f_n^D together with a unique pair q_n^D that are determined from

$$
\frac{d\pi_A}{df_A} = \left[v_A(q_A^D) - c_A - f_A^D\right]\frac{g(q_A^D)}{w} - \left[1 - G(q_A^D)\right] = 0,
$$
$$
\frac{d\pi_B}{df_B} = \left[v_B(q_B^D) - c_B - f_B^D\right]\frac{g(q_B^D)}{w} - G(q_B^D) = 0,
$$

so that we have that $f_A^D > 0$ and $f_B^D > 0$ in equilibrium. From (1) and (2) these values are uniquely determined. From substitution we have

$$
v_A(q_A^D) - c_A = w\left[\left(1 - q_A^D\right) + \frac{1 - G(q_A^D)}{g(q_A^D)}\right],
$$
$$
v_B(q_B^D) - c_B = w\left[q_B^D + \frac{G(q_B^D)}{g(q_B^D)}\right].
$$

We can proceed similarly for the case without disclosure. There, provided that the market again is not fully covered in equilibrium, we have a unique equilibrium with $f_n^{ND} > 0$ and q_n^{ND} determined by

$$
E[v_A(q) \mid q \geq q_A^{ND}] - c_A = w\left[\left(1 - q_A^{ND}\right) + \frac{1 - G(q_A^{ND})}{g(q_A^{ND})}\right], \tag{32}
$$
$$
E[v_B(q) \mid q < q_B^{ND}] - c_B = w\left[q_B^D + \frac{G(q_B^D)}{g(q_B^D)}\right].
$$

As long as the market is still not covered, it is immediate that sales of *both* products expand when w is reduced. The same holds also when commissions are not disclosed. These two observations highlight the key difference to the case of a fully covered market, in which a higher market share for one product comes at the expense of the share of the rival product. In the latter case, which firm's market share expands depended then on how a change (such as a shift in w or a change in the disclosure regime) affects the *relative* incentives of firms to pay commissions.

With respect to efficiency, it is also immediate that with disclosure sales of either product are too low in equilibrium, provided that the market is not fully covered. Instead, now the sales of *both* products can overshoot when w is low and commissions are not disclosed. To see this, suppose for simplicity that $c_A = c_B$. Observe that from (3) we obtain two efficient cutoffs $q_{A,FB}$ and $q_{B,FB}$ from the respective requirements that $v(q_{n,FB}) = c$, with $c_n = c$. These satisfy $0 < q_{B,FB} < 1/2 < q_{A,FB} < 1$. From (32) we have for low w that $q_B^{ND} > q_{B,FB}$ and $q_A^{ND} < q_{A,FB}$.

5. Information Example

Consider the two signal-generating distributions $F_A(s; a) = s^{a+1}$ and $F_B(s; a) = 1 - (1-s)^{a+1}$, where $a \geq 0$. By Bayes' rule the posterior belief is $q(s) = s^a/[s^a + (1-s)^a]$, with $q(0) = 0$ and $q(1) = 1$. Given that the unconditional distribution of the signal is $F(s) = 1/2 + [s^{a+1} - (1-s)^{a+1}]/2$, the distribution of the posterior in this power-signal example is

$$G(\widetilde{q}; a) = \frac{1}{2} + \frac{[q^{-1}(\widetilde{q})]^{a+1} - [1 - q^{-1}(\widetilde{q})]^{a+1}}{2}.$$

We now show that a induces a mean-preserving rotation in $G(\widetilde{q}; a)$, according to (18). To see that this is the case, note first that for $\widetilde{s} := q^{-1}(\widetilde{q})$ we obtain, after some transformations, $d\widetilde{s}/da = -\widetilde{s}(1-\widetilde{s})[\ln(\widetilde{s}) - \ln(1-\widetilde{s})]/a$, so that

$$\frac{dG(\widetilde{q}; a)}{da} = \frac{1}{2}\{[\widetilde{s}^{a+1}\ln(\widetilde{s}) - (1-\widetilde{s})^{a+1}\ln(1-\widetilde{s})] \tag{33}$$
$$-\frac{a+1}{a}(\ln(\widetilde{s}) - \ln(1-\widetilde{s}))[\widetilde{s}^{a+1}(1-\widetilde{s}) + \widetilde{s}(1-\widetilde{s})^{a+1}]\}.$$

Consider first $\widetilde{s} = 1/2$, for which it is straightforward that $dG(\widetilde{q}; a)/da = 0$. Next, for $\widetilde{s} < 1/2$ we can use from (33) that

$$\frac{dG(q; a)}{da} > \frac{1}{2}\{[\widetilde{s}^{a+1}\ln(\widetilde{s}) - (1-\widetilde{s})^{a+1}\ln(1-\widetilde{s})]$$
$$-(\ln(\widetilde{s}) - \ln(1-\widetilde{s}))[\widetilde{s}^{a+1}(1-\widetilde{s}) + \widetilde{s}(1-\widetilde{s})^{a+1}]\}$$
$$= \frac{1}{2}\{\underbrace{[\widetilde{s}\ln(\widetilde{s}) + (1-\widetilde{s})\ln(1-\widetilde{s})]}_{<0}\underbrace{[\widetilde{s}^{a+1} - (1-\widetilde{s})^{a+1}]}_{<0}\} > 0.$$

Similarly, for $\widetilde{s} > 1/2$ we have

$$\frac{dG(q; a)}{da} < \frac{1}{2}\{\underbrace{[\widetilde{s}\ln(\widetilde{s}) + (1-\widetilde{s})\ln(1-\widetilde{s})]}_{<0}\underbrace{[\widetilde{s}^{a+1} - (1-\widetilde{s})^{a+1}]}_{>0} < 0.$$

Finally, we show that $G(\widetilde{q}; a)$ also satisfies the hazard-rate property (1) for all $\widetilde{q}$ when $a \leq 1$. For this note that

$$\frac{d}{d\widetilde{q}}\frac{g(\widetilde{q}; a)}{1 - G(\widetilde{q}; a)}$$
$$= \frac{1+a}{2a}\frac{[\widetilde{s}^a + (1-\widetilde{s})^a]^2}{(1-\widetilde{s})^{a-1}\widetilde{s}^{a-1}}\left(\frac{a[1 - (1-\widetilde{s}^{a-1})(1 + (1-\widetilde{s})^{a-1})] + [\widetilde{s}^a + (1-\widetilde{s})^a]^2}{[1 - \widetilde{s}^{a+1} + (1-\widetilde{s})^{a+1}]^2}\right).$$

which is positive if $a[\widetilde{s}^{a-1} - (1-\widetilde{s}^{a-1})(1-\widetilde{s})^{a-1}] + [\widetilde{s}^a + (1-\widetilde{s})^a]^2 > 0$. For all $a < 1$, the first term on the left-hand side of this inequality is positive for all $\widetilde{s}$. When $a = 1$, the inequality holds strictly for all $\widetilde{s} > 0$ and weakly at $\widetilde{s} = 0$.

6. Covert Information Acquisition

Suppose that the quality of the adviser's information depends on a non-observable effort a exerted by the adviser at $t = 3$, i.e., after being matched with the customer. Consider the case with non-disclosure, in which the adviser's payoff is $u^{ND} - k(a)$, provided the recommendation is followed. The set of optimal values a depends only on the chosen cutoff $q^* = q^{ND}$, which is in turn independent of a. Importantly, the adviser's choice of effort thus does not depend directly on the level of commissions. In the symmetric case, where $c_A = c_B$ and $q^* = q^{ND} = q^D = 1/2$, this further implies that the effort chosen in $t = 3$ is independent of whether commissions are disclosed or not.

Continuing with the analysis, suppose for brevity that $k(a)$ is sufficiently convex so that the adviser's problem in $t = 3$ generates a unique solution. Depending on the choice of cutoff q^*, this level is uniquely determined from the first-order condition

$$w \left[\int_0^{q^*} \frac{dG(q; a)}{da} dq - \int_{q^*}^1 \frac{dG(q; a)}{da} dq \right] = k'(a)$$

and denoted by $a^*(q^*)$. By implicit differentiation we have that $da^*/dq^* = 0$ at $q^* = 1/2$. This implies that the firms' first-order conditions for the optimal choice of commissions evaluated at the symmetric outcome $f_n = f$ and $q^* = 1/2$, both with and without disclosure, do *not* change. Provided that a unique equilibrium in pure strategies exists both with and without disclosure, we can then fully apply the analysis from the main text by using $G(q) = G(q; a^*(1/2))$.

However, given that for non-marginal deviations from $f_n = f$ and thus $q^* = 1/2$ commissions affect the quality of information, the monotone hazard rate condition is no longer sufficient to ensure that the firms' program is strictly quasiconcave or that best responses intersect only once. See also the working paper version for an analysis when $q^D < 1/2$ and $q^{ND} < 1/2$ hold for the case of a monopoly firm.