Competition through Commissions and Kickbacks

By Roman Inderst and Marco Ottaviani*

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 intermediaries' 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 advisor responsive to supply-side incentives. We characterize situations when commonly adopted policies such as mandatory disclosure and caps on commissions have unintended welfare consequences.

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

---

1 Inderst: Johann Wolfgang Goethe University (IMFS) and Imperial College London: Mertonstrasse 17, 60054 Frankfurt am Main, Germany (e-mail: inderst@finance.uni-frankfurt.de); Ottaviani: Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208-2013 (e-mail: m-ottaviani@northwestern.edu). We are grateful for comments by Anne Coughlan, Simone Galperti, Umberto Garfagnini, Florian Hoffmann, Marcus Opp, Ron Siegel, four anonymous referees, and seminar participants at Bocconi, Gerzensee, IESE, Institute for Advanced Studies in Vienna, Kellogg School of Management, New York University, University of Bonn, University of Toronto, and University of Zurich. Earlier drafts of this paper circulated under the title "Intermediary Commissions and Kickbacks." Inderst gratefully acknowledges financial support from the European Research Council (FP7/ERC grant No. 229921).

2 To view additional materials, visit the article page at http://dx.doi.org/10.1257/aer.102.2.780.

1 Inducements may take the form of consultant fees, educational grants, royalties, funding for clinical trials, or travel grants. See 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 Hockenberry et al. (2011) for a descriptive analysis of the data on financial payment by orthopedic device makers to surgeons that became publicly available following the 2007 settlement between the leading manufacturers of joint implants and the Department of Justice.

2 A recent high-profile case was brought by former New York State Attorney General Eliot Spitzer against US insurance providers, most notably AIG. See Cummins and Doherty (2006) for a discussion and an empirical analysis of brokerage intermediation in the insurance market. See Jackson and Burlingame (2007) and Schwartz (2007) for a legal perspective on the pervasive use of commissions and kickbacks to compensate insurance, investment, real estate, and mortgage brokers.
believed to have led brokers to advise home buyers 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 advisor. The advisor 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 advisor is compensated through commissions paid by the firms. In addition, the advisor 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 have private information about which product best suits their needs, in our model customers are initially uninformed and must obtain this information from the advisor. Each firm is in a position to steer the advisor’s recommendation by stepping up its respective commission. The advisor 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 the classic model of price competition with horizontally differentiated products in Hotelling (1929). While in the Hotelling model the marginal customer reacts to the difference in products’ prices, in our model the advisor reacts to the difference in firms’ commissions. The advisor’s concern for suitability in our model plays the role of unit transportation cost and, thus, measures the inverse of the advisor’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 advisor’s suitability concern. Nevertheless, commissions are higher when the advisor 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 advisor. 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 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 nondisclosure scenario, as is surely the case when the advisor is highly concerned about suitability. Instead, when the advisor 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 advisors, and other intermediaries disclose to customers the commissions paid by product providers.4 Physicians and other healthcare 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 advisors.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 advisor, but also by the quality of the advisor's information. For example, advisors with superior training are in a position to obtain better information about suitability. We show how the quality of the advisor'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 advisor can extract a larger fraction of customers' additional benefits

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. 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 United Kingdom, 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 beginning in 2012 prevents financial advisors in the United Kingdom 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.
from the improved advice. This is because, without 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 advisor's incentives to invest in information.

The paper proceeds by presenting our contribution to the literature in Section I. Section II formulates the model. Section III characterizes the baseline scenario with undisclosed commissions, while Section IV analyzes the regime with disclosure. Section IV compares the two disclosure regimes from a welfare perspective. Sections V to VII endogenize the advisor's concern for suitability through fines, the threat of losing the franchise, and professional standards. Section VIII endogenizes information acquisition. Section IX concludes. The main Appendix at the end of the text collects the proofs. An online Appendix in six sections discusses robustness and extensions.

I. Contribution to Literature

There is a dearth of literature on commissions and kickbacks. 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. 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 nonstrategically 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 kickbacks. 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 advisor can credibly communicate to the customer. In our equilibrium model, the customer responds rationally to the advisor'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 the classic model of competition between two price-setting firms in Hotelling (1929). While the middlemen of Biglaiser (1993) and certification intermediaries of Lizzeri (1999) provide vertical information about quality, our information intermediaries provide horizontal information about match suitability. Thus, our model relates to work by Lewis and Sappington (1994); Moscarini and Ottaviani (2001); Johnson and Myatt (2006); Ganuza and Penalva (2010); and Bar-Isaac, Caruana, and Cuñat (2010) on sellers' incentives to provide information to customers. In our model, however, this information is provided by an intermediary advisor rather than directly by the sellers. It is only through commissions that the competing sellers can affect the information the advisor conveys to the customer.

While common agency models following Bernheim and 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 advisor who in
turn communicates with a customer. In a model in which brokers steer customers to appropriate providers, Colwell and 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 advisor’s concern for suitability.

The intermediary’s incentives to provide biased advice are influenced by firms’ commissions, rather than being specified exogenously as in Crawford and Sobel (1982) and in most of the theoretical literature on strategic communication. Morgan and Stocken (2003) analyze communication by a sender with uncertain bias, while Li and Madarász (2008) characterize how disclosure of such an exogenous and uncertain bias affects the resulting cheap-talk equilibrium. Once the advisor’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 advisor is concerned about suitability, Durbin and Iyer (2009) allow a (single) biased principal to influence the preferences of an advisor who aims to be perceived as being incorruptible. 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 advisor as an intermediary is also a key difference from Bolton, Freixas, and 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. 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 sales-related bonus pay. Because there the seller (rather than the advisor, 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 advisor 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.

II. 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 \( v_{\theta} \) if the product matches the state and utility \( v_{\neg \theta} \) 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 advisor’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 VIII we further parametrize the quality of the advisor’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 advisor, 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,

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

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

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

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

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

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 advisor and sell directly to the customer. To ensure that either product can be sold with advice, we further stipulate that

$$
\int_{1/2}^1 v_A(q) \frac{g(q)}{1 - G(1/2)} dq = \int_0^{1/2} v_B(q) \frac{g(q)}{G(1/2)} dq > c_B.
$$
Hence, if the advisor 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 advisor cares directly about whether the purchased product is suitable for the customer’s needs. We capture the advisor’s concern for suitability in a flexible way by positing that the advisor derives utility $w_h$ when the customer purchases the more suitable product, and utility $w_l$ otherwise. The advisor’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 advisor 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 advisor when intermediating a purchase generates sufficiently large benefits from other simultaneous or subsequent transactions with the same customer. We discuss below (and analyze in online Appendix 4) the case in which the advisor sometimes wants to recommend to customers not to purchase any product.

In the paper we develop three foundations for the advisor’s concern for suitability:

- The advisor 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. This case is analyzed in detail in Section V, where we also derive policy implications.
- The advisor’s preference for suitability may arise from reputational concerns or fear of losing future business prospects when the business license is revoked by an authority. To this end, Section VI shows that our results carry over to a dynamic model in which the suitability concern $w$ depends on the expected future commissions the advisor risks losing.
- The advisor may be motivated by a professional concern for the customer’s well-being. In this case, we stipulate that the advisor places some weight $\gamma$ on suitability: $w_l = \gamma v_l$ and $w_h = \gamma v_h$, so that $w = \gamma(v_h - v_l)$. This specification is analyzed in Section VII.

---

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 Keest (2009) and “Should Auto Dealers Avoid New Regulation?” in 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 (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 (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
Timing.—To influence the advisor, at period $t = 1$ firms simultaneously set their respective commissions (or fees) $f_n$ to be paid to the advisor conditional on the sale of their product. The advisor 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 advisor 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 advisor 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 fails, we rule out payments from or to customers that are contingent on the realized utility, $v_j$ or $v_h$. At period $t = 3$, the advisor 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 advisor.

III. 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 advisor’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. Following 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 the wary beliefs of McAfee and Schwartz (1994). It is convenient to stipulate passive beliefs about commissions also with respect to the advisor’s recommendation.
in $t = 3$, so that in any pure-strategy equilibrium the customer holds (point) beliefs $f_n$ 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 $w_0 < w_1$ we can restrict consideration to only two messages for the advisor, which correspond to the two products, $A$ and $B$. Ignoring the payoff-equivalent outcome in which the messages are swapped, in equilibrium the advisor’s recommendation is followed by the customer. Then, the advisor’s expected payoff from recommending product $A$ equals $f_A + q w_h + (1 - q)w_1$ and from recommending product $B$ equals $f_B + (1 - q)w_h + q w_1$. When both products are recommended with positive probability, the advisor recommends $A$ rather than $B$ when $q > q^*$, with the cutoff given by

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

Specifying that the advisor 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 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 $v_A (\tilde{q}) - p_A = v_B (\tilde{q}) - p_B$, so that the marginal customer type is $\tilde{q} = 1/2 - (p_B - p_A)/(2(v_h - v_l))$. 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 advisor’s concern 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 advisor 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 advisor’s expected cutoff, denoted by $\hat{q}^*$, which is obtained by plugging the expected commissions $\hat{f}_n$ into (5). Again, if substitution of $\hat{f}_n$ into (5) does not lead to an interior threshold, we set $\hat{q}^* = 0$ or $\hat{q}^* = 1$, 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, $q^\ast$, the customer’s conditional expected valuation for each product is given by

$$P_A(q^\ast) = \int_{q^\ast}^1 v_A(q) \frac{g(q)}{1 - G(q)} \, dq \equiv E[v_A(q) | q \geq q^\ast],$$

$$P_B(q^\ast) = \int_0^{q^\ast} v_B(q) \frac{g(q)}{G(q)} \, dq \equiv E[v_B(q) | q < q^\ast].$$

Setting a price $p_n < P_n(q^\ast)$ 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 $q^\ast$; 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 advisor’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

$$\pi_A = [p_A - f_A - c_A][1 - G(q^\ast)],$$

$$\pi_B = [p_B - f_B - c_B]G(q^\ast).$$

Profits depend directly on the actual cutoff $q^\ast$, 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 advisor’s recommendation, the marginal customer type $q^\ast$ is determined by the advisor’s indifference condition (5). In addition, given that firms optimally set $p_n = P_n(q^\ast)$ according to (6), profits depend on the expected cutoff $q^\ast$ and, thus, on the expected commissions that are anticipated by the customer.

Differentiating firms’ profits, we obtain firms’ best responses for given $q^\ast$ as follows. For firm $A$, we have

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

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

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

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 advisor's recommendation to changes in commissions is given by $\left| \frac{dq^*}{df_\text{A}} \right| = \frac{dq^*}{df_\text{B}} = \frac{1}{2w}$. 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 (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 $f_n$, while customer expectations must be satisfied: $f_n = f_n$ for $n = A, B$ and thus, $q^* = q^*$. 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 $f_\text{ND}, p_\text{ND},$ and $q_\text{ND}$.

When the advisor 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 advisor 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 $q_\text{ND} = 1/2$. When firms have the same incentives to steer advice, competition thus creates balanced incentives for the advisor, irrespective of the extent of the advisor'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 $q_\text{ND} < 1/2$ results in equilibrium. Note that because an increase of a firm’s commission must also be paid inframarginally (i.e., for all $q > q_\text{ND}$ for firm A and for all $q < q_\text{ND}$ 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 advisor’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

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

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 advisor’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.

**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 advisor 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:

\[
(11) \quad \pi_A^{ND} = 2w \frac{[1 - G(q^{ND})]^2}{g(q^{ND})},
\]

\[
\pi_B^{ND} = 2w \frac{[G(q^{ND})]^2}{g(q^{ND})}.
\]

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. This need not, however, 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 from (3) that trade cannot occur without advice. We used this restriction to rule out the possibility that a firm can deviate profitably by sufficiently undercutting the rival and, thereby, induce the customer to buy its product even against the advisor’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 advisor could essentially prescribe a particular product, thereby leaving the customer with only 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 advisor realizes the payoffs \( w_l \) and \( w_h \) only when the respective product is purchased through the intermediated channel, the advisor 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 advisor, 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 advisor’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 advisor’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 (see online Appendix 3 for further analysis).

IV. Disclosure of Commissions

With disclosed commissions, the customer can directly infer the advisor’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 expression (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, \( q^\ast \). 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 conditional valuations, we obtain the best responses with disclosure:

\[
\begin{align*}
    f_A &= v_A(q^*) - c_A - 2w \frac{1 - G(q^*)}{g(q^*)}, \\
    f_B &= v_B(q^*) - c_B - 2w \frac{G(q^*)}{g(q^*)}.
\end{align*}
\]

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 \(f^D, p^D, \) 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 advisor’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

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

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 advisor 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 advisor, \( q^{ND} = q^D = 1/2 \). 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:

\[
E[v_A(q) | q > q^*] - v_A(q^*) > E[v_B(q) | q < q^*] - v_B(q^*),
\]

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 advisor’s belief is symmetric around \( q = 1/2 \). At \( q^* = 0 \), instead, this condition holds strictly, because then the left-hand side equals \( (\nu_l + \nu_h)/2 - \nu_l \), 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 guarantee that \( E[v_A(q) | q > q^*] - v_A(q^*) \) is everywhere strictly decreasing and that \( E[v_B(q) | q < q^*] - v_B(q^*) \) 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 advisor 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 \( A \).
V. Welfare and Policy

Turning to welfare analysis, in this section we specify that the advisor’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 \(\omega\) is strictly quasiconcave in \(q^*\) and maximized at \(q^* = q_{FB}\) with

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

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 advisor’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 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 the more cost-efficient 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 advisor always leads to efficient advice. Consider the asymmetric case with \(c_A < c_B\). With disclosure, firms internalize how their commissions affect the advisor’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 advisor’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.

\(^{10}\)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, Poser and Fanto (2007).
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 advisor 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 advisor’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 advisor’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 advisor’s recommendation has on the quality of advice. In the presence of full disclosure, the imposition of penalties on the advisor 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 advisor; however, policymakers often do not have complete control over disclosure.\(^{11}\) Third, even though regulations that

---

\(^{11}\)Full disclosure of commissions is often difficult to implement in practice because firms have many opportunities to influence the advisor’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).
impose penalties and mandate disclosure should be optimized jointly, in many markets policymakers do not have complete control of penalties either, for example because these penalties and the advisor’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 advisor prefers that the customer would sometimes not purchase. As noted above, by specifying that \( w_1 > w_0 \) 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. In the more general case in which the advisor, 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 \( f_{\pi} \leq \tilde{f} \), where \( \tilde{f} \) 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 \( q^{\text{ND}} = 1/2 \) 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 \( q^{\text{ND}} \) 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 \( \tilde{f}_{FB} > 0 \) 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.

**VI. Franchise Value and Reputation**

Rather than imposing fines, a regulator or supervisor could also reduce the value of the advisor’s franchise by revoking the business or professional license. To analyze this possibility we now embed our model in a streamlined dynamic setting. At the beginning of each period, \( \tau = 0, 1, \ldots \), 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 advisor discounts profits by the factor \( \delta < 1 \). When \( v = v_p \), we posit that the advisor loses the license and is then replaced by another advisor, who then has the same access to customers. This setting ensures that revoking a license 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 advisor for more than one sale
(however, see online Appendix 2), we stipulate that each period the customers have different needs and their choices are 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 advisor’s expected utility is given by

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

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 III and IV, 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 advisor earns. Denote the respective values without and with disclosure by $U^{ND}$ and $U^D$. Interestingly, when expected future commissions are higher, the advisor’s recommendation becomes less sensitive to commissions because more is at stake when the advisor loses the license and, thus, the franchise value.

**PROPOSITION 5:** Suppose that the advisor’s concern for suitability arises endogenously from the risk of losing the franchise following the realization of $v = v_\delta$, given that the license is then revoked. When future profits become relatively unimportant (lower $\delta$), 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 nondisclosure when $\delta < \delta_D$, while disclosure reduces efficiency when $\delta > \delta_D$.

When $\delta$ is low and the advisor is not very motivated to give suitable advice because future profits and, thus, the value of the advisor’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 $\delta$ is high and the advisor 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 advisor 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 more heavily relative to future profits (low $\delta$) 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.\(^\text{12}\)

\(^{12}\)A case in point is the liberalization of the UK pension market at the end of the 1980s, which led to a ramp-up of commissions and ensuing allegations of unsuitable advice and egregious mis-selling in the 1990s. As also stressed
While so far we have specified that the advisor loses the franchise following the revocation of the license 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 advisor in the market. Unless \( v_h \) is observed, customers expect to receive a perfectly uninformative recommendation from the advisor in the future, and the advisor no longer expects to make any profits from commissions in the market. In this way, we can also guarantee that the advisor always prefers to recommend a purchase rather than no purchase at all.

VII. Professional Concern

Even when we abstract from liability or the threat of losing future business, concern for suitability arises when the advisor 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 \( \gamma \) on the respective realizations, we have \( w_l = \gamma v_h \), \( w_h = \gamma v_h \), and \( w = \gamma (v_h - v_l) \). When we take this into account for welfare, efficiency is now maximized when the advisor’s cutoff is equal to

\[
\tilde{q}_{FB} = \frac{1}{2} - \frac{c_B - c_A}{2(1 + \gamma)(v_h - v_l)}.
\]

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

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 \( \tilde{q}_{FB} \) in (17) results, this wedge is no longer present. We still, however, have \( q^D > \tilde{q}_{FB} \) because of the agency problem between firms and the advisor, 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 advisor’s concern for suitability arises from placing weight \( \gamma \) on the suitability of the customer’s choice. Disclosure of commissions increases efficiency when \( \gamma \) is low, but decreases efficiency when \( \gamma \) is high.

Proposition 6 shows that the results from Corollary 1 and Proposition 5 are also robust when the advisor’s concern for suitability arises only from professional standards. While we obtain in Corollary 1 and Proposition 5 a unique cutoff on the
penalty \( w \) or the importance of future profits \( \delta \), however, 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 \( \gamma \).

VIII. 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 advisor’s information. To endogenize the advisor’s information, consider a family of distribution functions \( G(q; a) \), where \( a \in [a, \bar{a}] \) 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

\[
\frac{dG(q; a)}{da} \geq 0 \quad \text{for} \quad q \leq \frac{1}{2}
\]

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 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 advisor’s information, rather than on the advisor’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 advisor’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 \( k'(a) = 0 \) 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 advisor is matched with a particular customer.

The advisor’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.

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

\[
\pi^{ND} = \frac{w}{2g(1/2; \alpha)}.
\]

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

\[
\pi^D = \pi^{ND} + \left[ E[v_A(q) | q \geq q^D] - v_A(q^D) \right] [1 - G(q^D)].
\]

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 advisor becomes better informed.

When commissions are not disclosed, firms fully compete away, through higher commissions, any increase in customers’ valuation that arises when the advisor 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 \( g(1/2; \alpha) \) is smaller, i.e., when it is less likely that the advisor 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 advisor 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.

Advisor Incentives.—The advisor’s expected payoff when commissions are not disclosed is

\[
u^{ND} = w_l + \int_0^{1/2} [f^{ND}_B - wq] g(q; \alpha) dq + \int_{1/2}^1 [f^{ND}_A - w(1 - q)] g(q; \alpha) dq - k(\alpha).
\]

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

\[
u^{ND} = (w_l + f^{ND}) + w \left[ \int_0^{1/2} G(q; \alpha) dq - \int_{1/2}^1 G(q; \alpha) dq \right] - k(\alpha).
\]
We can obtain an analogous expression for the advisor’s resulting payoff with disclosure

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

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

When the information quality \( a \) increases, there is clearly a direct positive effect on the payoff the advisor derives from the concern for suitability. This is captured by the fact that the second term in (22) in brackets increases strictly by (18). In addition, \( a \) has an indirect effect on the advisor’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 advisor 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 advisor fully extracts the benefits that better information quality has for customers.

Taking Section V’s efficiency criterion for the case in which the advisor’s concern for suitability arises from regulatory penalties, we have:

**PROPOSITION 7:** Suppose that the advisor 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 advisor’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 advisor 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 advisor’s concern for suitability. When \( w \) is sufficiently large, however, the advisor’s choice of \( a \) becomes excessively high.

Note finally that the effect that a change in the quality of the advisor’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 advisor 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 advisor's choice of \( a \). Note from expressions (21) and (22) that commissions do not directly affect the impact that information quality has on the advisor'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 advisor 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.

**IX. 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 advisor is little concerned for suitability because in that case the cost-efficient firm’s market share is too high without disclosure. When, instead, the advisor’s suitability concern is sufficiently strong, disclosure reduces welfare. Furthermore, disclosure reduces the advisor’s incentives to acquire information in advance.

Some recent experimental studies suggest that imposing mandatory disclosure of commissions may have additional drawbacks. Lacko and Pappalardo (2007) conjecture that disclosed commissions may prevent information-overloaded customers from adequately digesting other payoff-relevant facts. In another experimental study, Cain, Loewenstein, and Moore (2005) argue that disclosure of bias may lead advisors 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 advisors have adjusted their expectations through repeated experience.

In our model, a single advisor obtained information that was sufficient to judge the suitability of competing products. Alternatively, one could imagine that different advisors 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 advisors would become private information for the customer, with repercussions also on the equilibrium communication strategy of the first advisor.

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 advisor—but we leave a formal analysis of relational commissions to future
research. 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.

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 \( \hat{q}^* \), \( f_A \) is increasing and \( f_B \) is decreasing, so that from (5) we obtain a nonincreasing and continuous function \( q^*(\hat{q}^*) \). 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 \( \hat{q}^* = q^* \), 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

\[
\frac{d\pi_B}{df_B}\bigg|_{f_B=0} = [P_B(\hat{q}^*) - c_B] \frac{g(q^*)}{2w} - G(q^*).
\]

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

\[
w < \tilde{w} \equiv g(1/2)[E[v_B(q) | q < 1/2] - c_B].
\]

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

\[
\frac{d\pi_A}{df_A}\bigg|_{f_A=0} = [P_A(\hat{q}^*) - c_A] \frac{g(q^*)}{2w} - [1 - G(q^*)].
\]

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 \( w \geq w' = g(1/2)[E[v_A(q) | q > 1/2] - c_A] \). Since \( c_A \leq c_B \) and \( E[v_A(q) | q > 1/2] = E[v_B(q) | q < 1/2] \) by (4), we have \( w' \geq \tilde{w} \). In this case, \( q^{ND} \) is characterized by (10).

---

14From 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 USC. § 1320a-7b(b).
Finally, a threshold for \( w \) that also provides 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 \( q^* = q^* \). 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 \( E[v_A(q) | q > 1/2] = E[v_B(q) | q < 1/2] \).

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 \( f_B^{ND} \) 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

\[
\frac{df_A}{df_B} = \frac{v'_A(q^*) - 2w \frac{d}{dq} \frac{1 - G(q^*)}{g(q^*)}}{2w + v'_A(q^*) - 2w \frac{d}{dq} \frac{1 - G(q^*)}{g(q^*)}} > 0;
\]

\[
\frac{df_B}{df_A} = \frac{v'_B(q^*) - 2w \frac{d}{dq} \frac{G(q^*)}{g(q^*)}}{-2w + v'_B(q^*) - 2w \frac{d}{dq} \frac{G(q^*)}{g(q^*)}} > 0.
\]

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 \( f_A^{D} > 0 \) is that

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

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 \( w < w' = g(1/2)[v_A(1/2) - c_A] \) follows from strict quasiconcavity of \( \pi_A \) 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 $f^D_n$ follow from the same arguments as those for $q^{ND}$ and $f^{ND}_n$ 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 $E[v_B(q) | q < q^{ND}] > v_B(q^D)$, we have that $f^{ND}_B > f^D_B$. When $q^{ND} > q^D$, instead, a symmetric argument implies $f^{ND}_A > f^A$. Next, we show that indeed both $f^{ND}_n > f^n$ when strictly positive. To see this, suppose $f^{ND}_n > f^n$, as we already established that this must hold for at least one firm. Thus, arguing to a contradiction, $f^{ND}_n < f^n$ would then imply $q^{ND} < q^D$, but this would imply, from comparing (9) with (12), also the opposite, namely, $f^{ND}_n > f^A$. In case $f^{ND}_n > f^D_B$, we can contradict $f^{ND}_n < f^n$ 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 $E[v_A(q) | q \geq q^{*}] - v_A(q^{*})$ is everywhere strictly decreasing, while $E[v_B(q) | q < q^{*}] - v_B(q^{*})$ 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

$\int_0^{q^{*}} \frac{G(q)}{G(q^{*})} dq.$

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 Theorem 1 in Bagnoli and Bergstrom (2005), 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 $E[v_B(q) | q < q_L] - E[v_A(q) | q \geq q_L] = c_B - c_A$. 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 $\omega^{ND}$ and $\omega^D$, and welfare at the first-best cutoff by $\omega^{FB}$. 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 \( q^{ND} \rightarrow q^*_{FB} < q_{FB} \) as \( w \rightarrow 0 \). 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 advisor optimally chooses \( q^* \), as characterized by (5) with \( w = U \). Each period, the two firms maximize \( \pi_A \) and \( \pi_B \). 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 advisor, 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 \( \delta \) increases. But as \( w = U \), we know from Proposition 2 that then \( f^{ND}_A \) and \( f^{ND}_B \) must both increase, which from (16), given that \( q^{ND} \) is chosen optimally by the advisor, 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 \( \delta^* > 0 \) for \( \delta \) so that \( q^{ND} < q_{FB} \) for higher \( \delta \) and \( q^{ND} > q_{FB} \) for lower \( \delta \) follows then from Corollary 1 and observing that \( U = w \rightarrow 0 \) as \( \delta \rightarrow 0 \).

PROOF OF PROPOSITION 6:
Given that an increase in \( \gamma \) 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 \( \gamma \). This follows from Propositions 1 and 2. Also, we then have \( q^D > q^{ND} \) and \( q^D > \bar{q}_{FB} \), albeit now \( \bar{q}_{FB} \) also strictly increases in \( w \) and thus in \( \gamma \). It remains to determine when \( q^{ND} > \bar{q}_{FB} \) and when the opposite holds.

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

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

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 \( \bar{q}_{FB} \rightarrow q_{FB} < 1/2 \) as \( w \rightarrow 0 \). When \( w \) is sufficiently large, the converse holds strictly, so that \( q^{ND} > \bar{q}_{FB} \).

PROOF OF PROPOSITION 7:
We consider the advisor’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

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

where we use $c_n = c$. From comparison with $u^{ND}$ we have $du^{ND}/da < d\tilde{\omega}/da$ for all sufficiently low $w$. Note that then $d\tilde{\omega}/da < 0$ 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**


