ADVISORS WITH HIDDEN MOTIVES*

Paula Onuchic
University of Oxford

April, 2024

Abstract

A seller discloses evidence about an object to a potential buyer, who doesn't know the object's value or the profitability of its sale (the seller's motives). I characterize optimal disclosure rules that balance two goals: maximizing the overall probability of sale, and steering sales from lower- to higher-profitability objects. I consider a policy that reveals the seller's motives to the buyer, and show that its effectiveness in inducing the seller to disclose evidence depends on the curvature of the buyer's demand for the object. This result refines our understanding of effective regulation of advisor-advisee communication with and without commitment.

1 Introduction

People frequently take advice from advisors with hidden motives: broker-dealers and other financial advisors counsel investors, but also receive sales commissions from financial product providers; digital influencers provide product reviews to their followers, but these are often sponsored content; doctors inform patients of the effectiveness of different drugs and procedures, but may be rewarded by pharmaceutical companies; magazines and newspapers selectively publish pieces of reporting that align with their editorial bias. In all the mentioned settings, information receivers understand that information providers may be biased, but are not fully aware of the extent of the conflict. For example, in the context of financial brokers, clients understand that brokers receive sales commissions from some product providers, but may not know the size of the commissions on each product; in the social media context, followers understand that influencers post sponsored content, but cannot inherently know which publications are paid advertising.

Such conflicts of interest between advisors and advisees receive much regulatory attention. The most common recommendation, and often instituted regulation, is *transparency*, prescribing advisors to let advisees know the real interests behind their recommendations. (Think of the SEC mandating that

*I am grateful for guidance from Debraj Ray, Erik Madsen and Ariel Rubinstein, and for long discussions with Joshua Weiss and Samuel Kapon. I also thank Margaret Meyer, David Pearce, Ludvig Sinander, Dilip Abreu, Nageeb Ali, Ricardo Alonso, Arjada Bardhi, Heski Bar-Isaac, Sylvain Chassang, Navin Kartik, Laurent Mathevet, Luis Rayo, Mauricio Ribeiro, and Dezsö Szalay for their helpful comments.

financial advisors disclose commissions they receive from financial product providers, or the FTC and the UK CMA mandating digital influencers to mark their posts as "paid content" when they are sponsored by brands, and take other steps to clarify their relation with product producers to their content consumers.[1])

In this paper, I propose a model where a seller commits to a policy to disclose evidence about an object's value to a buyer who then chooses whether to acquire the object. The seller is an advisor with *hidden motives*, because the buyer does not know how profitable the object's sale is to the seller, and is therefore unable to assess precisely how interested the seller is in pushing the sale. In this context, I characterize the seller's optimal disclosure policies, and show that they are chosen to balance two potentially conflicting goals: to maximize the overall probability of the object's sale, and to steer sales from lower- to higher-profitability objects. Having characterized optimal advice, I consider the introduction of a transparency policy that reveals to the buyer the profitability of the object's sale. I show that as a response to such policy, the seller does not necessarily become more willing to disclose information about the object to the buyer. Specifically, the effectiveness of the transparency policy is linked to the curvature of the buyer's demand for the object. This novel observation is in contrast with results about the effectiveness of transparency policies under communication protocols without commitment.

In the model, prior to the realization of the object's profitability and of a signal about its value, the seller commits to a rule to disclose signal realizations to the buyer. This rule assigns to each realization of the signal a probability that the realization is disclosed to the buyer; and, importantly, this rule may also depend on the profitability of the object's sale. The buyer is Bayesian and updates their belief about the object's value based on any information the seller reveals and on the seller's disclosure policy itself. The probability that the buyer purchases the object is given by their "demand function," which is increasing in their posterior expectation of the object's value. This "demand function" is taken as a primitive of the model, but in Section 2.1, I provide possible micro-foundations that see the demand function as arising from the buyer choosing between the object being offered by the seller and one or more alternative outside options. The proposed micro-foundation introduces a possible interpretation of the curvature of the demand function as a measure of competition in the market to which the seller belongs. Specifically, a "more convex" demand function would arise in a more competitive market, where the buyer has access to more potential outside options.

In Theorem 1, I show that optimal disclosure rules follow a threshold structure. Each signal realization about the object's value is classified as either "good news," if above some endogenously determined threshold, or "bad news." For each good news realization, there is a *profitability threshold* such that the realization is disclosed if the object's profitability is above the threshold, and concealed otherwise. Conversely, each piece of bad news is concealed if the object is sufficiently profitable, and revealed

[1] In Section 2.2, I discuss the market for financial advice as a "leading application" and interpret different features of the model in that context. The market for social media influencers is equally a good fit. Because of the recent prominence of the influencer market, official regulations have ben implemented only in the past few years; for example, the FTC revised its "Guides Concerning Use of Endorsements and Testimonials in Advertising" in 2023, partly to put forth new guidance on which are necessary steps for influencers to clearly reveal their material connections to brands. See Ershov and Mitchell (2023), who study the effects of advertising disclosure regulations in social media markets; and Ershov, He, and Seiler (2023), who quantify the prevalence of undisclosed influencer posts on Twitter.

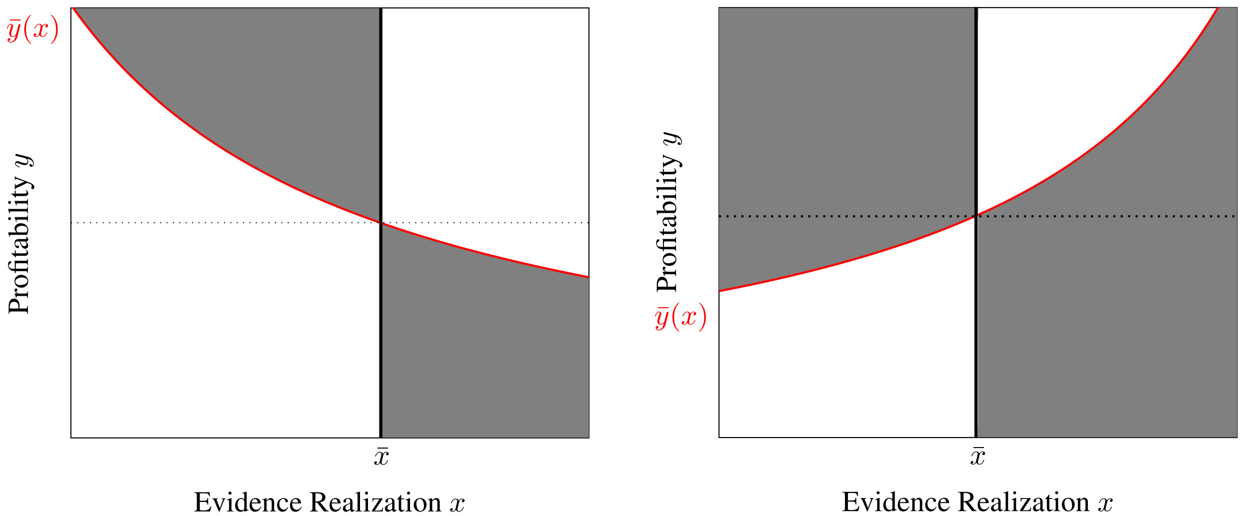

Figure 1: Optimal disclosure rule when the buyer's demand function is **convex** (left panel) and **concave** (right panel). These images correspond to the optimal disclosure rules corresponding to the examples in sections 3.2.1 and 3.2.2, respectively. The gray areas represent evidence realizations that are optimally concealed by the seller, and the white areas are optimally disclosed.

otherwise. Although this threshold-structure description of optimal signals is simple, it encompasses disclosure policies with distinct properties, which are determined by the shape of the profitability threshold function. Theorem 1 also characterizes optimal profitability thresholds, and Corollaries 1 and 2 show that their shape depends on the curvature of the buyer's demand function.

For an illustration, see Figure 1. In each panel, the x-axis represents the expected value of the object induced by a realization of the evidence, and the y-axis represents the object's profitability. The threshold $\bar{x}$ divides good-news and bad-news evidence realizations; and the function $\bar{y}(x)$ is the optimal profitability threshold dividing disclosure and non-disclosure regions. The left panel of Figure 1 illustrates optimal disclosure when the buyer's demand for the object is convex. In that case, the seller can maximize the overall probability of sale by disclosing all evidence realizations; but optimally chooses to conceal some realizations in order to steer the buyer from low- to high-profitability objects. As a consequence, extreme evidence realizations (very bad news or very good news) are always disclosed, but some bad news are concealed when the object's sale is very profitable, as are some good news if the object is less profitable. This description is summarized by the optimal profitability threshold being a *decreasing* function of the value of the evidence realization. Conversely, if the buyer's demand for the object is concave, then the optimal profitability threshold is an *increasing* function — as depicted in the right panel of Figure 1. When demand is concave, this shape is a result of balancing the seller's desire to maximize overall sales by concealing all realizations; and to steer sales towards high-profitability objects by selectively disclosing some evidence.

Having characterized optimal disclosure for a seller with hidden motives, I then consider a policy intervention in the spirit of commonly used transparency policies. In the model, introducing a transparency policy means imposing that on top of observing any information about the object that is willingly disclosed by the seller, the buyer also sees the object's profitability to the seller. Therefore, under the transparency policy, the seller cannot use strategic non-disclosure in order "pool" signal realizations across objects with different profitability levels. The results in section 4 compare the seller's optimal disclosure policy with and without mandated transparency.

Proposition 2 shows that if the buyer's demand function is concave, then the seller voluntarily discloses more evidence to the buyer when their motives are hidden. The intuition is that under hidden motives some selective information disclosure allows them to steer sales from low- to high-profitability objects. And indeed in order to effectively steer sales, the seller *must* disclose some information. If otherwise the seller's motives are revealed to the buyer, such steering through selective disclosure becomes impossible; and instead the seller optimally conceals all information so as to maximize the overall probability that an object (of any profitability) is sold. The opposite result holds when the buyer's demand function is convex; and therefore transparency policies are effective in such demand environment.[2] In terms of a regulatory takeaway, Proposition 2 argues that the effectiveness of a transparency policy can depend on the curvature of the buyer's demand function, highlighting that there is no "one size fits all" policy to optimally regulate advice markets. Rather, it is important to fit the regulation to specific features of each market. Using the interpretation introduced in Section 2.1, which relates the curvature of the buyer's demand function to the degree of competition in the relevant market, we learn that mandating transparency is an effective policy in very competitive markets, but may not be so in markets where the buyer does not have many alternatives to the object being offered by the seller.

While Proposition 2 compares optimal disclosure policies with and without a mandated transparency regulation in cases where the buyer's demand function is convex or concave, it is silent about the effectiveness of such regulation when the demand function has other shapes. Propositions 3 and 4 provide complementary results regarding more general specifications for the buyer's demand function. Specifically, Proposition 3 relates the local curvature of the demand function to a local increase or decrease in disclosure due to the regulation. And Proposition 4 shows that, when the seller's motives are hidden, their ability to steer the buyer leads them to increase the disclosure of "good news" when the object's sale is highly profitable and of "bad news" for less profitable objects.

In section 5, I examine two versions of the model, in which I vary the communication protocol relative to the "disclosure with commitment" benchmark. First, I consider a variation in which the seller makes disclosure choices *without commitment*. That is, they first observe the realization of the signal regarding the object's value, as well as the object's profitability, and only then choose whether to disclose the evidence to the buyer. The main observation is that transparency policies are always effective in such an environment, independently of the shape of the buyer's demand function. This result is in line with an informal intuition that is often behind the introduction of transparency policies, namely that if an advisee knows the intention behind the information they receive, then they can take the advice

[2] If the demand function is convex and the seller has hidden motives, then they selectively conceal information from the buyer (bad news about high-profitability objects and good news about low-profitability objects), so as to steer the buyer from low- to high-profitability sales. If the seller's motives are revealed by the mandatory transparency policy, then steering through selective concealment becomes impossible; and instead the seller optimally discloses all information so as to maximize the overall probability that an object (of any profitability) is sold.

with a "healthy skepticism" which incentivizes the advisor to provide informative advice. Although this intuition is justified by the no commitment protocol, there are reasons (such as reputation-building in repeated interactions) why a degree of "commitment" exists in the relationship between the advisor and the advisee. In that light, the current paper shows that if there is some degree of commitment in that relationship, then transparency of the advisor's motives is not always an effective or sufficient method to regulate the advice market. Finally, in Section 5.2, I briefly comment on a variation in which the seller can commit to signaling structures more general than simple disclosure policies.

1.1 Related Literature

This paper contributes to the literature on disclosure of verifiable information, started with Grossman (1981) and Milgrom (1981). For a survey on that literature, see Milgrom (2008). In particular, in my model the sender uses "simple evidence" as in Dye (1985) and Jung and Kwon (1988), in which a piece of verifiable "evidence" is either entirely disclosed to the receiver or concealed. My model mainly departs from that literature by considering a disclosure problem where the sender has hidden motives, and can commit to a disclosure rule prior to the realization of the evidence.

In assuming that the sender can commit ex-ante to a signaling technology, my paper relates to the large literature on information design mainly stemming from Kamenica and Gentzkow (2011) and Rayo and Segal (2010). My paper, like Rayo and Segal's (2010) work, departs from most of that literature in that it considers a problem of *multidimensional* information design. And it differs from Rayo and Segal (2010) both because I consider a disclosure communication protocol, which allows me to characterize optimal policies when the buyer's demand function is not linear,[3] and because I study the introduction of policies that make the seller's motives transparent.[4]

Also within information design, my paper contributes to a growing literature studying *constrained information design*. These are problems where the sender is subject to additional constraints, beyond Bayesian plausibility, when choosing a signal structure. For example, Mensch (2021) and Onuchic ⓡ Ray (2022) consider monotonicity constraints. See Doval and Skreta (2024) and references therein for other examples. This paper considers a sender who is constrained to disclosure strategies.

Some previous literature studies environments with endogenous information acquisition, and argues that policies that ostensibly improve information disclosure can actually depress information acquisition and decrease the sender's informativeness. Matthews and Postlewaite (1985) argue that by imposing that the seller disclose any evidence they may have about the quality of an object, a regulator can remove their incentives to test the object's quality in the first place. Libgober (2022) shows that a receiver may prefer to commit to not knowing some dimensions of an experiment produced by a sender, so as to incentivize

[3]Most of Rayo and Segal's (2010) characterization results apply to the linear specification. With nonlinear demand functions, the optimal disclosure rule in my model sometimes "pools ordered prospects" – in Rayo and Segal's (2010) language, ordered prospects are two objects whose values and profitabilities are ordered in the same way. For an example, see the optimal disclosure rule depicted in the right panel of Figure 1. One of Rayo and Segal's (2010) main characterization results is that, in the linear benchmark, optimal signals never pool ordered prospects.

[4]The term "transparent motives" was coined by Lipnowski and Ravid (2020) to refer to the sender's preferences being state-independent. In my model, under mandated transparency, the seller acts *as if* they have transparent motives, in that sense.

information acquisition by a sender. Kartik, Lee and Suen (2017) show that an advisee may prefer to solicit advice from just one biased expert even when others (of equal or opposite bias) are available, because in the presence of more advisors, each individual expert free rides on the information acquired by the other experts. Che and Kartik (2009) study an environment in which a sender and receiver share the same preferences, buthold different priors over the distribution of states. They show that the sender may invest more in acquiring information when the disagreement between their priors is larger.[5] In contrast with this literature, my paper argues that mandated transparency of the seller's motives affect the advisor's incentives to reveal exogenous information to the buyer.

Finally, this paper also relates to previous work studying cheap talk models in which the sender has hidden motives. For example, Sobel (1985), Morris (2001), and Morgan and Stocken (2003), study environments with cheap talk communication in which the receiver does not know whether the sender's preferences are aligned with their own. My paper is most related to Li and Madarasz (2008), which studies a version of Crawford and Sobel's (1982) cheap talk environment, with the additional assumption that the receiver does not know the size or direction of the sender's bias. Like me, Li and Madarasz (2008) ask whether instituting a policy that mandates the transparency of the sender's motives (their bias in that case) is gainful for the receiver. Also like me, they find that such mandated transparency policy may be ineffective. Our papers differ in that we consider distinct communication protocols, and highlight different features of the environment that determine the effectiveness of mandated transparency regulation (the curvature of the buyer's demand function, and the degree of commitment in the advisor-advisee relationship, in my case).

2 Environment

A seller wishes to sell an object to a buyer who does not know their value for the object, or how profitable the object's sale is to the seller. The object's *value* to the buyer, denoted $x \in \mathcal{X} = [x_{min}, x_{max}]$, and its *profitability* to the seller, denoted $y \in \mathcal{Y} = [y_{min}, y_{max}]$, with $y_{min} \geqslant 0$, are drawn from a joint distribution commonly known by seller and buyer.

Before the buyer decides whether to purchase the object, the seller can disclose to them a signal conveying some information about the object's value (see below for details on the communication protocol). After observing any conveyed information, the buyer forms a Bayesian posterior belief about the object's value, with some expected value $\hat{x} \in \mathcal{X}$. Given their posterior $\hat{x}$ about the object's value, the buyer purchases the object with probability $p(\hat{x})$, where $p : \mathcal{X} \to [0, 1]$ is a strictly increasing and continuously differentiable "demand function." If the object is purchased, the seller receives a payoff equal to the object's profitability, y. Otherwise, the seller's payoff is 0.

[5] See also Shishkin (2023) and DeMarzo, Kremer, and Skrzypacz (2019) about information acquisition by the sender in a Dye (1985) framework. And see Szalay (2005) and Ball and Gao (2024) about information acquisition by a biased agent in a delegation context. An earlier version of this paper also shows that the introduction of a transparency policy can hinder the seller's incentives to acquire information about the objects value; and therefore also affect the informativeness of their advice through that channel.

Information Disclosure Protocol. The seller has access to a signal that conveys information about the object's value. The signal, $\pi : \mathcal{X} \times \mathcal{Y} \to \Delta\mathcal{M}$, is a measurable function mapping the object's value and profitability to a distribution of messages in a measurable set of possible messages, $\mathcal{M}$. Given signal π, for each signal realization $m \in \mathcal{M}$, $\hat{x}(m) = \mathbb{E}(x|m)$ is the implied Bayesian posterior expected value of the object. I assume that, given a signal realization m, knowing the profitability of the object conveys no additional information about the object's value; that is, $\mathbb{E}(x|m, y) = \mathbb{E}(x|m) = \hat{x}(m)$.[6] In other words, if an agent observes message m from signal π, they interpret it as "the object's expected value is $\hat{x}(m)$," independently of their belief about the object's profitability.

Given this independence, I refer to a signal realization and to the posterior mean it induces interchangeably; and equate the signal π with its induced distribution of posterior means. Formally, let F be the joint distribution over $\mathcal{X} \times \mathcal{Y}$ of signal realizations and the object's profitability, implied by the prior distribution of values and profitabilities and by the signal π. Denote by F_Y the marginal profitability distribution, and by $F_{X|y}$ the distribution signal realizations conditional on a profitability $y \in \mathcal{Y}$. I assume that distributions F_Y and $F_{X|y}$, for each $y \in \mathcal{Y}$, have strictly positive densities f_Y and $f_{X|y}$ for each y.

At an initial stage, the seller commits to a rule to disclose signal realizations to the buyer. A disclosure rule is a measurable map from a signal realization *and the object's profitability* into a probability that the realization is disclosed, $d : \mathcal{X} \times \mathcal{Y} \to [0, 1]$. Note that the disclosure decision depends both on the signal realization and on the profitability of the object, so in practice the seller commits to rules to disclose signal realizations, which are conditional on the object's profitability.

If signal realization $\hat{x}$ is disclosed, the buyer's posterior mean after observing it is, by definition, exactly $\hat{x}$. If otherwise the signal realization is not disclosed, the buyer's posterior mean is computed using Bayes' Rule, accounting for the disclosure rule. Formally, if $\int_{\mathcal{Y}} \int_{\mathcal{X}} [1 - d(x, y)]\, dF_{X|y}(x) dF_Y(y) > 0$,[7]

$$x^{ND} = \frac{\int_{\mathcal{Y}} \int_{\mathcal{X}} x\,(1 - d(x, y))\, dF_{X|y}(x) dF_Y(y)}{\int_{\mathcal{Y}} \int_{\mathcal{X}} (1 - d(x, y))\, dF_{X|y}(x) dF_Y(y)}, \tag{1}$$

which is the expected value conditional on non-disclosure. The average object profitability given that a realization is not disclosed is analogously given by

$$y^{ND} = \frac{\int_{\mathcal{Y}} \int_{\mathcal{X}} y\,(1 - d(x, y))\, dF_{X|y}(x) dF_Y(y)}{\int_{\mathcal{Y}} \int_{\mathcal{X}} (1 - d(x, y))\, dF_{X|y}(x) dF_Y(y)}.$$

2.1 Micro-Foundations for the "Demand Function" p

Throughout this paper, I regard the buyer as a passive agent, a "receiver" who sees (or does not see) information about the object, forms a belief about the object's expected value, and buys it or not according to some exogenously given "demand function" p. A possible micro-foundation for this demand function,

[6] This condition holds, for example, if the object's value and its profitability are independent and the signal π is independent of the object's profitability. Alternatively, it holds if the signal fully reveal's the object's value, in which case no additional information can be conveyed by the knowledge of the object's profitability.

[7] Otherwise, non-disclosure is "off-path," and we fix x^{ND} at some value in $\mathcal{X}$.

which follows the description in Rayo and Segal (2010), is given below.

Privately-known outside option. A risk-neutral buyer chooses between acquiring the object being sold by the seller or taking an outside option. An example of an outside option would be of buying another object somewhere else, or refraining from buying altogether. The value of the buyer's outside option x_o is private information, distributed according to F_o. Once the buyer sees all the provided information about the object's value, and forms belief $\hat{x}$, they purchase the object if $\hat{x} > x_o$ and do not purchase it otherwise. From the seller's perspective, a purchase then happens with probability $F_o(\hat{x})$, the probability that the expected value of the object is greater than that of the outside option. In this case, the "demand function" p coincides with the distribution of outside option values F_o.[8] The assumption that p is increasing and continuously differentiable requires then that the cdf F_o be continuously differentiable.

Competing sellers. A different micro-foundation, in the spirit of Hwang, Kim, and Boleslavsky (2023), describes a buyer who has access to the object being offered by the seller and to various potential outside options, perhaps referring to objects being sold by alternative sellers. The value of the product sold by each of the outside sellers is unknown to the inside seller, but each is distributed according to F_o, and value draws are independent across inside and each of the outside objects. Therefore, if there are n outside sellers, the best outside option available to the buyer, $\max(x_o)$ is distributed according to $\max(x_o) \sim (F_o)^n$. The buyer purchases the inside object then if the expected value of the inside object, $\hat{x}$, is greater than $\max(x_o)$. This happens with probability $(F_o)^n(\hat{x})$. In this case, the "demand function" p therefore coincides with $(F_o)^n$. Again, the assumption that p is increasing and continuously differentiable requires then that the cdf F_o be continuously differentiable.

Note moreover, that an *increase in competitiveness*, in the sense of an increase in the number of competitors n, maps into an increase in the convexity of the demand function. Formally, if $n' > n$, then demand function $p' = (F_o)^{n'}$ is more convex than demand function $p = (F_o)^n$, because the former function is a strictly increasing and convex transformation of the latter. And moreover, if n is large enough, then the demand function $p = (F_o)^n$ is a convex function in the entirety of its support.

The results stated in the paper show that the characterization of optimal disclosure rules, and policy implications in terms of transparency mandates, depend on the curvature of the demand function. Once these results are stated, I refer back to this interpretation of the convexity of the demand function as related to the competitiveness in the market to which the seller belongs.

2.2 Financial Advisors with Hidden Motives

The features of the model just introduced can be interpreted in the context of a brokerage company who provides financial advice to investors, but also receives commissions for the sale of financial products.

[8]Alternatively, one may think that the seller communicates with a population of possible buyers, and each buyer decides between making a purchase from the seller and taking their own private outside option. In that case, F_o is the distribution of outside options in that population of buyers, and $p = F_o$ describes the amount of sales to be made by the seller when they induce a certain posterior belief about the object's value on the population of buyers.

In that scenario, clients understand that brokers receive sales commissions from some product sales, and therefore wish to sell products often. However, in each instance when the advisor suggests a financial product, the client may not know the exact size of the commissions on that particular product (and hence the advisor's motives are hidden).

Investors commonly receive advice in the form of reports about financial products. In addition to publicly available data about companies and industries, a report on a particular product includes forecasts and evaluations produced by in-house research teams. In line with the model, a research report about a financial product is a piece of evidence which induces on the investor who sees it some perception of the product's value. Most research departments are subject to "Chinese wall" regulations which institute a separation between research teams and client-facing sales teams within the same institution. These regulations aim to ensure that the analyses produced by a research teams are not biased by the companies' profit-seeking interests. Accordingly, in the model, the evidence about the object's value is not biased. Rather, the seller can only affect the buyer's behavior by selectively choosing which pieces of (unbiased) evidence to reveal and which to strategically conceal.

The exercise in section 4 studies a regulation which reveals the seller's motives (the profitability of the object) to the buyer. Because investors often rely on professional advice when making investment decisions, financial advisory is a highly regulated field. (And has become increasingly so since the Great Recession and the institution of the Dodd-Frank Act.) In the United States, the regulation varies across states, and also across categories within the field – for example, financial advisors and insurance agents are often subject to more stringent regulation than broker-dealers, who often also provide advice. Specific details notwithstanding, much of the regulation focuses on making the interests of financial advisors transparent to investors who consult them. For example, they might be required to disclose other sources of compensation received beyond service fees, professional affiliations with another broker-dealer or securities issuer, and other potential or existing conflicts of interest. Other types of regulation also forbid advisors from receiving some types of compensation. For instance, the UK and the Netherlands have altogether imposed bans on commission payments for some types of financial advisors.

Commitment to Information Disclosure. An important feature of the model is that the seller can *commit* to an evidence-disclosure rule prior to observing the evidence realization. A common interpretation in the information design literature sees the commitment assumption as a stand in for a sender's desire to build and sustain credibility with a receiver in a game of repeated interaction. Think of a circumstance where a financial advisor repeatedly interacts with an investor, informing them about products' values over a long time. And suppose that this investor follows the advisor's recommendation, but also sees the outcome of their investment decisions — so they can tell, to some extent, whether the advice was sound. Over time, the investor can "punish" unsatisfactory advice from the seller with future incredulity – maybe by not following their recommendation, maybe by seeking another advisor. Through that mechanism, in the repeated game, the advisor's optimal disclosure strategy would be such that the buyer's interpretation of a "no disclosure" message coincides with the interpretation yielded in the one-shot game under the commitment assumption. A version of this argument is made formally made by Best and Quigley (2023), and a similar argument linking commitment and reputation is provided by Mathevet, Pearce and Stacchetti (2022).[9] The literature also provides other justifications for the "commitment assumption" — for example, Deb, Pai, and Said (2023) show that the seller could implement the commitment signaling strategy by communicating with the buyer through an informed third party. In that case, commitment to an employment contract with that agent is a stand in for commitment to the communication strategy.

Of course, commitment is not an assumption that perfectly reflects the interaction between an investor and their financial advisor. But neither is the assumption that the advisor has no commitment at all — truth is somewhere in between. The analysis in this paper highlights that optimal disclosure and the effectiveness of transparency policies differs significantly under a commitment protocol, as compared to a protocol with no commitment (discussed in section 5). Consequently, one of the contributions of this paper is to refine our understanding of effective regulation in advice markets by highlighting how they vary depending on the degree of commitment in the advisor-advisee relation.

Literature on Financial Advisors with Hidden Motives. In a series of papers in 2012, Inderst and Ottaviani (2012.1, 2012.2, 2012.3) propose models of brokers and financial advisors compensated through commissions. Competing sellers offer commissions to an advisor, knowing that they will steer business to sellers that offer higher compensation. The authors show that biased commissioned advisors may benefit welfare by steering business to high-commission firms who are also more cost efficient at the expense of providing information to buyers. In my model, the sender's distribution of profitability is given, and I show that hidden motives can improve welfare precisely by increasing the amount of information provided to the consumer.

There is also a large literature that studies the provision of information by Credit Rating Agencies that are financed by fees paid by issuers of financial products. Some key papers in this literature are Bolton, Freixas, and Shapiro (2012), Opp, Opp and Harris (2013), Bar-Isaac and Shapiro (2012) and Skreta and Veldkamp (2009). In this literature, the CRA receives payments equally from all issuers of financial products. In my model, the main concern is that the advisor benefits some products over others because they have different profitabilities.

[9] Best and Quigley (2023) find that reputation in a repeated interaction between sender and receiver is *generally not* a substitute for commitment if the sender's optimal strategy (with commitment) involves mixing between messages after some realization of the state. In that case, for the sender to be willing to randomize (in the repeated game), the sender must be made indifferent between messages; Best and Quigley (2023) show that to implement such indifference, a non-negligible amount of continuation surplus must be "burnt," even in the patient limit. However, in the disclosure context studied in this paper, we will see that the sender's optimal strategy does not involve randomization, and therefore reputation can be a substitute for commitment.

Further, note that in the "repeated game" version of the current environment with hidden motives, it would not be necessary for the receiver to observe the profitability of the object in order to establish the equivalence between reputation and commitment. Because the buyer only cares about the object's value in this context, it would suffice for them to observe the object's value in order to "punish" unsatisfactory advice.

3 Optimal Disclosure

Suppose the seller commits to a disclosure rule d. The probability that the object is sold, conditional on its profitability being y is

$$P(y,d) = \int_{\mathcal{X}} \left[d(x,y)p(x) + (1-d(x,y))p(x^{ND})\right] dF_{X|y}(x). \tag{2}$$

To understand (2), first remember that $F_{X|y}$ is the distribution of signal realizations given that the object has profitability y. Suppose a signal realization x is disclosed, which happens with probability $d(x,y)$. Then the object is sold with probability $p(x)$, which is reflected in the first term inside the integral of (2). As for the second term, with probability $1-d(x,y)$ the realization x is not disclosed. In that case, the object is sold with probability $p(x^{ND})$, where x^{ND} is as given in (1).

The seller's expected payoff from committing to disclosure rule d is then

$$\Pi(d) = \mathbb{E}\left[yP(y,d)\right] = \mathbb{E}(y)\mathbb{E}\left[P(y,d)\right] + \text{Cov}\left[y, P(y,d)\right]. \tag{3}$$

In (3), I split the seller's payoff into two terms, expressing that the seller's objective can be seen as twofold. According to the first term, the seller wishes to maximize the overall expected probability of sale, which is multiplied by the average profitability. Per the second term, they seek to maximize the covariance between the object's profitability and its probability of sale. This covariance term reflects the seller's desire to steer the buyer from purchasing low-profitability objects to purchasing those with high profitability. These two objectives are sometimes at odds, and the characterization provided below illustrates how optimal disclosure balances the two goals.

The first result, Theorem 1, provides a *threshold characterization* of disclosure rules that maximize the seller's value. There is a threshold value $\bar{x}$ such that each signal realization is classified as either good news, if x is larger than $\bar{x}$, or bad news, if $x < \bar{x}$. For a good news realization $x > \bar{x}$, there is a profitability threshold $\bar{y}(x)$ such that a signal realization is disclosed if and only if the object's profitability is above that threshold. Conversely, each bad news realization $x < \bar{x}$ is disclosed if and only if the object's profitability is below the threshold $\bar{y}(x)$.

Theorem 1. *An optimal disclosure rule exists and every optimal rule d^* has a threshold structure: There is a threshold value $\bar{x} \in \mathcal{X}$ and a threshold profitability $\bar{y} : \mathcal{X} \to \mathcal{Y}$ such that d^* almost everywhere satisfies $d^*(x,y) \in \{0,1\}$ and*

$$d^*(x,y) = 1 \Leftrightarrow (x-\bar{x})(y-\bar{y}(x)) \geqslant 0. \tag{4}$$

The threshold value satisfies $\bar{x} = x^{ND}$, and, for $x \neq x^{ND}$, the threshold profitability $\bar{y}(x)$ satisfies

$$\bar{y}(x) = y^{ND}\left[\frac{p'(x^{ND})(x^{ND}-x)}{p(x^{ND})-p(x)}\right]. \tag{5}$$

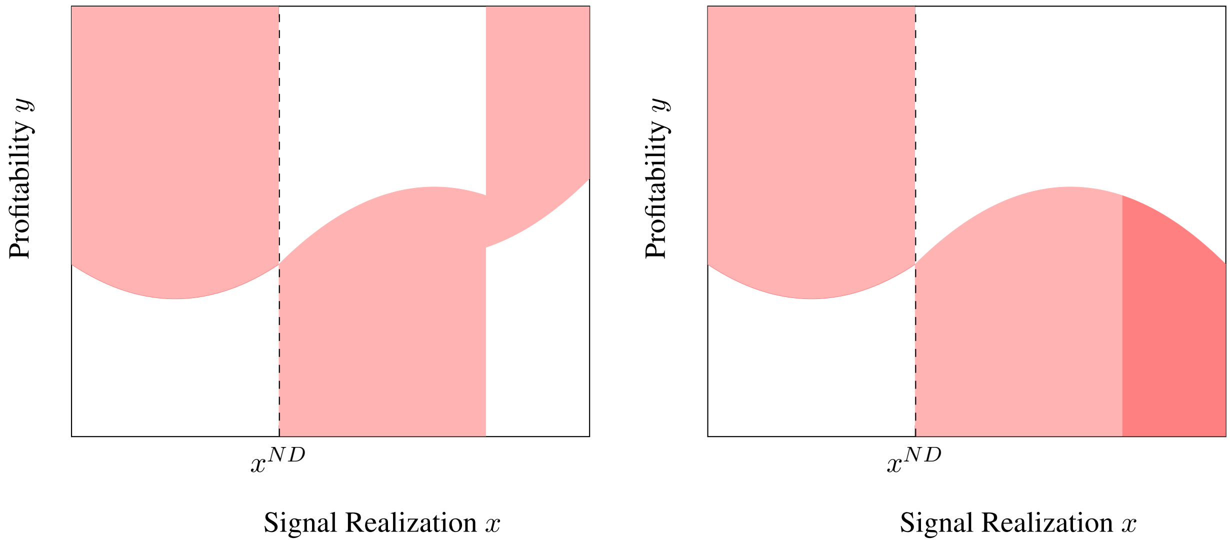

Figure 2: Building disclosure rule $\hat{d}$ from disclosure rule d: in each panel, the colored areas representing zones of no disclosure and the white areas representing zones of disclosure. The left-hand side panel illustrates a possible disclosure rule d that does not have a "threshold structure." The right-hand side panel depicts a disclosure rule $\hat{d}$, which has a threshold structure, and is derived from d according to (6), (7), and (8).

A complete proof of Theorem 1, including the characterization of the threshold profitability function $\bar{y}(\cdot)$ is provided in the Appendix. I now provide an intuitive discussion, arguing that any disclosure rule which leads to no disclosure with positive probability and does not satisfy the threshold structure described in (4) can be improved upon by a rule that does satisfy (4). Start with one such disclosure rule d according to which no disclosure happens with positive probability, and which does not satisfy the threshold structure. (For example, the disclosure rule depicted in the left-hand side panel of Figure2.) Let x^{ND} be its implied non-disclosure posterior mean. Define then an alternative rule, $\hat{d}$, that discloses each realization x with the same probability as d, but has a threshold structure. That is, if $x \leqslant x^{ND}$, let

$$\hat{d}(x, y) = \begin{cases} 1, \text{ if } y \leqslant \hat{y}(x) \\ 0, \text{ if } y > \hat{y}(x) \end{cases} \tag{6}$$

and if $x > x^{ND}$, let

$$\hat{d}(x, y) = \begin{cases} 0, \text{ if } y < \hat{y}(x) \\ 1, \text{ if } y \geqslant \hat{y}(x) \end{cases} \tag{7}$$

where the thresholds $\hat{y}$ are calibrated such that, for each realization x,

$$\int_{\mathcal{Y}} \hat{d}(x, y) dF_{Y|x}(y) = \int_{\mathcal{Y}} d(x, y) dF_{Y|x}(y), \tag{8}$$

where $F_{Y|x}$ is the profitability distribution conditional on a signal realization x. The right-hand side panel of Figure 2 illustrates the "improvement" $\hat{d}$ derived from d on the left-hand side panel, according to (6), (7), and (8). Condition (8) implies that d and $\hat{d}$ induce the same x^{ND}, and so $\hat{d}$ satisfies (4), with $\bar{x} = x^{ND}$.

By moving from d to $\hat{d}$, the seller shifts the disclosure probability of bad news to low profitability objects and of good news to high profitability objects, while maintaining the distribution of posterior mean values that is induced on the buyer. It is easy to see (and I argue formally in the Appendix), that: *Claim 1.* d and $\hat{d}$ produce the same overall probability of sale, because the distribution of posterior mean values is unchanged; and *Claim 2.* $\hat{d}$ induces a strictly larger covariance between sales and profitability than d, because the change increases the probability that very profitable objects are sold, and decreases that probability for less profitable objects. These facts imply that $\hat{d}$ yields a strictly larger expected seller payoff than d, as desired.

The value and profitability thresholds described in Theorem 1 partition the value-profitability space into four "quadrants" — see, for example, Figures 1 and 3, which illustrate the quadrants defined by the optimal disclosure rules. The first and third quadrants represent regions where the seller and buyer have aligned interests, either because both value and profitability are high or because both value and profitability are low. Signal realizations in these "alignment" regions are optimally disclosed to the buyer. Conversely, the second and fourth quadrants represent areas of misalignment between advisor and advisee, and therefore these signal realizations are optimally concealed.

Interior vs. Corner Optimal Disclosure Rules. Theorem 1 shows that an optimal disclosure rule is such that, for some $\hat{x} \in \mathcal{X}$ and $\hat{y} \in \mathcal{Y}$,

$$(x - \hat{x})(y - \bar{y}(x)) \geqslant 0 \Rightarrow d_{\hat{x},\hat{y}}(x, y) = 1, \quad (x - \bar{x})(y - \bar{y}(x)) < 0 \Rightarrow d_{\hat{x},\hat{y}}(x, y) = 0,$$

$$\text{where } \bar{y}(x) = \hat{y}\left[\frac{p'(\hat{x})(\hat{x} - x)}{p(\hat{x}) - p(x)}\right].$$

And the following system of two equations must hold:

$$\begin{cases} \hat{x} = \mathbb{E}\left[x \middle| \text{no disclosure region implied by } d_{\hat{x},\hat{y}}\right] \\ \hat{y} = \mathbb{E}\left[y \middle| \text{no disclosure region implied by } d_{\hat{x},\hat{y}}\right] \end{cases} \tag{9}$$

But note that, if $\hat{x} = \inf(\mathcal{X})$ and $\hat{y} = \inf(\mathcal{Y})$, or $\hat{x} = \sup(\mathcal{X})$ and $\hat{y} = \sup(\mathcal{Y})$, it is possible for the no disclosure region implied by $d_{\hat{x},\hat{y}}$ to be empty.[10] In that case, the expected values on the right hand side of the equations in (9) are not defined by Bayesian updating. For convenience, I take the stance that the buyer's "off-path" beliefs of no disclosure are such that (9) is vacuously satisfied. This means that, if the no disclosure set is empty at one or both corners — $\hat{x} = \inf(\mathcal{X})$ and $\hat{y} = \inf(\mathcal{Y})$, or $\hat{x} = \sup(\mathcal{X})$ and $\hat{y} = \sup(\mathcal{Y})$ — then these "corner solutions" with full disclosure are candidate optimal disclosure rules.

[10]For any other "corner cases," in which $\hat{x} \in \{\inf(\mathcal{X}), \sup(\mathcal{X})\}$ or $\hat{y} \in \{\inf(\mathcal{Y}), \sup(\mathcal{Y})\}$, the implied disclosure rule $d_{\hat{x},\hat{y}}$ is necessarily such that no disclosure happens with positive probability. In the proof of Lemma 1, I argue that these corner disclosure rules cannot be solutions to the seller's problem, as they do not satisfy system (9).

Lemma 1 and Proposition 1 below delineate conditions so that the optimal disclosure rule does not involve full disclosure and does not correspond to one of the two potential corner solutions. Specifically, Lemma 1 shows that if system (9) has an interior solution, then such solution must imply a disclosure rule that yields strictly higher payoff to the seller than full disclosure.

Next, Proposition 1 proposes three conditions that guarantee that system (9) has an interior solution. The first condition guarantees that, for any $\hat{x} \in \mathcal{X}$ and $\hat{y} \in \mathcal{Y}$, the implied disclosure rule $d_{\hat{x},\hat{y}}$ induces no disclosure with positive probability; in that case, I show that (9) must have an interior solution. This first condition is satisfied, for example, if the demand function p is strictly concave; but it can also hold for non-concave demand functions. The second condition states that (9) also has an interior solution if the demand function p is affine. Finally, condition 3 shows that the same holds if p is "close enough" to being an affine function; specifically, this condition guarantees that the optimal disclosure rule is interior for strictly convex demand functions that are "not too convex." (In section 3.2.1, I work out an example where the demand function p is convex and the optimal disclosure rule is interior.)

Lemma 1. *If system (9) has an interior solution, with $\hat{x} \in int(\mathcal{X})$ and $\hat{y} \in int(\mathcal{Y})$, then the optimal disclosure rule is interior and such that no disclosure happens with positive probability.*

Proposition 1.

1. *If there exist $x, x' \in \mathcal{X}$ such that*

$$\frac{p'(\inf(\mathcal{X}))(x - \inf(\mathcal{X}))}{p(x) - p(\inf(\mathcal{X}))} > 1 \text{ and } \frac{p'(\sup(\mathcal{X}))(x' - \sup(\mathcal{X}))}{p(x') - p(\sup(\mathcal{X}))} < 1,$$

   then the optimal disclosure rule is interior and no disclosure happens with positive probability.

2. *If p is an affine function, then the optimal disclosure rule is interior and no disclosure happens with positive probability.*

3. *If $p = \alpha p_1 + (1-\alpha)p_2$, where p_1 is an affine function, then there exists $\bar{\alpha} \in (0,1)$ such that if $\alpha > \bar{\alpha}$, the optimal disclosure rule is interior and no disclosure happens with positive probability.*

Steering and Credibility of Optimal Disclosure The characterization of optimal disclosure rules highlights the steering logic behind optimal advice in real financial advice markets. On the one hand, advisors understand that withholding evidence makes advisor-receivers skeptical — the literature on evidence disclosure has empirically documented the link between non-disclosure and skepticism. On the other hand, by committing to a threshold disclosure rule, an advisor can induce an ambiguous meaning on the advisee's skepticism: they cannot fully disentangle whether "no news" means "bad news about the object's value (skepticism about the object's quality), or whether "no news" means "the object's sale is not profitable to the seller" (skepticism about the object's profitability). By creating such ambiguity, the advisor can profitably steer the advisee. Indeed, it is well documented in financial advice markets that brokers conceal bad news about companies in which they have financial interests and that financial advisors' recommendations favor products with high commissions.[11]

In this model, steering is made possible because the seller is able to commit to a disclosure rule, and can therefore affect the buyer's interpretation of "no disclosure" by pooling together high-profitability-low-value realizations with low-profitability-high-value realizations. But is it reasonable to expect such commitment power from the advisor? As mentioned in section 2.2, commitment can be seen as a stand-in for reputation in a repeated interaction between advisor and advisee. Beyond this argument, a recent paper by Lin and Liu (2023) proposes a notion of credibility for information design problems: A disclosure policy is credible if the sender cannot profit from tampering with her messages while keeping the message distribution unchanged. Their reasoning is that if a sender "deviates" from an information policy in a way that keeps the marginal probability of sending each message unchanged, then that deviation would be "undetectable"; and if a policy is such that there are "undetectable" deviations that would benefit the sender, then such policy is not credible.

According to this definition, the optimal disclosure rules described in Theorem 1 are credible. To see, note that any disclosure rule other than the optimal rule as given in Theorem 1, but which induces that same marginal distribution over messages, must invariably swap the disclosure of a realization (x, y) and a different realization (x, y'); that is, two states with the same evidence realization x. But as per the argument in the proof of Theorem 1, if one such "undetectable swap" were available and were indeed profitable, then the starting disclosure rule must not have been optimal in the first place.

3.1 Linear Demand.

Sections 3.1 and 3.2 now provide further characterization of optimal disclosure under different assumptions about the buyer's "demand function" p. Corollary 1 applies Theorem 1 when p is affine.

Corollary 1. *If the demand function p is affine, then the thresholds defining an optimal disclosure rule satisfy $\bar{x} = x^{ND} \in int\,(\mathcal{X})$ and $\bar{y}(x) = y^{ND} \in int\,(\mathcal{Y})$ for all $x \in \mathcal{X}$.*

When the demand function p faced by the seller is affine, all disclosure rules yield the same overall probability of sale. That is, for any two disclosure rules d and d', $\mathbb{E}\,[P(y, d)] = \mathbb{E}\,[P(y, d')]$. This fact is an implication of the martingale property of posterior beliefs: we know that the expected posterior belief of the buyer about the value of the object has to equal the buyer's prior belief about the object's value. Because the probability of sale is an affine function of such posterior belief, the expected probability of sale has to equal the "ex-ante probability fo sale."

Consequently, in choosing a disclosure policy, there is no scope for the seller to increase or decrease the overall probability that the buyer purchases the object. Rather, the seller solely distributes this constant sale probability between high and low profitability objects. In order to optimally steer sales from low- to high-profitability objects, the seller uses a constant threshold $\bar{y}(x) = y^{ND}$, thereby effectively

[11] See, for example, Anagol, Cole and Sarkar (2017) and Eckardt and Rathke-Doppner (2010) about insurance brokers; Chalmers and Reuter (2020) on retirement plans; Inderst and Ottaviani (2012.2) on general financial advice. For a survey on quality disclosure and certification, see Dranove and Jin (2010).

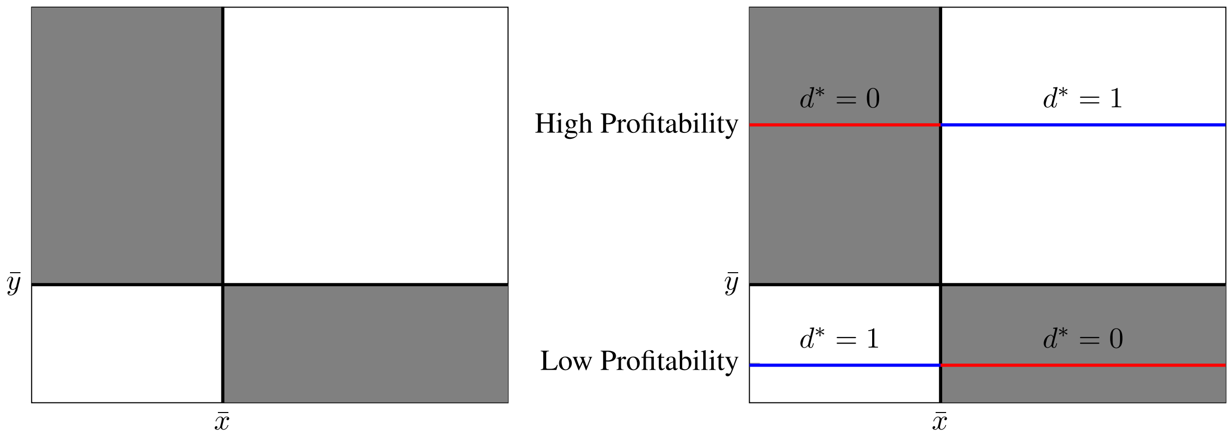

Figure 3: Left-hand panel: optimal disclosure rule when p is **affine**. The gray areas represent signal realizations that are optimally concealed by the seller, and white areas are optimally disclosed. The right-hand panel illustrates that, in this case, the object is classified as either a high-profitability object or a low-profitability object. For the former, signal realizations are disclosed if and only if they are "good news," whereas for the latter signal realizations are disclosed if and only if they represent "bad news."

assigning the object to one of two classes high profitability, with $y > y^{ND}$, and low profitability, with $y < y^{ND}$. Such an optimal disclosure rule is depicted in Figure 3; its right-hand panel illustrates that, for high-profitability objects, signal realizations are disclosed if and only if they represent "good news," while the opposite is true for low-profitability objects.

3.2 Nonlinear Demand.

If the demand function p is not affine, then the amount of information disclosed by the seller impacts the overall probability of sale. Observation 1 below states that increasing the amount of information about the object's value that is disclosed to the buyer increases the overall probability that the object is sold if the demand function is convex, and decreases it if the demand function is concave. To formally state this result, we say that a disclosure rule d has *more disclosure* than a disclosure rule d' if $d(x, y) \geqslant d'(x, y)$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$; strictly so if this inequality is strict for a subset of $\mathcal{X} \times \mathcal{Y}$ with positive measure.

Observation 1. *Suppose d has more disclosure than d'. Then if p is strictly convex (concave), d yields a higher (lower) sale probability than d'; and strictly so if d has strictly more disclosure than d'.*

With a nonlinear demand function, the seller transfers sales from low- to high-profitability objects *at the expense of* the total probability that the object is sold. If the seller faces a convex demand function, he wishes to disclose more information in order to maximize the sale probability; but maximizing the covariance between profitability and probability of sale requires the concealment of some realizations. Conversely, if the demand function is concave, overall sale probability increases when information is concealed from the buyer, but to improve the covariance between sales and profitability, the seller must disclose some signal realizations. The following corollary describes features of optimal disclosure in these cases.

Corollary 2. *If p is strictly convex (concave), an optimal disclosure rule has a strictly decreasing (increasing) profitability threshold function $\bar{y}(x)$, satisfying $\bar{y}(x^{ND}) = y^{ND}$.*

An optimal disclosure rule when p is strictly convex is represented in the left panel of Figure 1. Unlike in the linear demand case, some "very bad news" may be optimally disclosed even when the object is highly profitable; and, conversely, some "very good news" are disclosed to the receiver even when the object has low profitability. This feature is derived from the tension between the desire to maximize the probability of sale and to covary the probability of sales with profitability. In the opposite case, where p is strictly concave, due to an analogous tension, the sender optimally conceals some "very bad news" about low-profitability objects and some "very good news" about high-profitability objects. This is depicted in the right panel of Figure 1.

For illustration, I now provide two worked-out examples — corresponding to the cases depicted in Figure 1 — and fully describe the respective optimal disclosure rules.

3.2.1 Example: Optimal Disclosure under Convex Demand Function

Suppose an object's value and it's profitability are independently distributed, each distributed according to the uniform distribution $U[0,1]$. Moreover, signal realizations are fully revealing, so that $F_Y = U[0,1]$ and $F_{X|y} = U[0,1]$ for each $y \in [0,1]$. Further, let the demand function be given by $p(x) = x^2$.

From Theorem 1, we know that the thresholds defining the optimal disclosure rule should satisfy:

$$\bar{x} = x^{ND}, \text{ and}$$

$$\bar{y}(x) = y^{ND}\left[\frac{2x^{ND}(x^{ND}-x)}{(x^{ND})^2 - x}\right] = y^{ND}\left[\frac{2x^{ND}}{x^{ND}+x}\right].$$

Moreover, given these thresholds, x^{ND} and y^{ND} must indeed correspond to the Bayesian posteriors implied by no disclosure: $x^{ND} = \mathbb{E}(x|\text{no disclosure})$, and $y^{ND} = \mathbb{E}(y|\text{no disclosure})$. Numerically, I find that there is a unique pair $(x^{ND}, y^{ND}) \in (0,1)^2$ that satisfies these conditions. The pair is given by

$$x^{ND} = 0.5824 \text{ and } y^{ND} = 0.5098.$$

Because there is a unique such pair, we know that it must correspond to the thresholds in the optimal disclosure rule. For an illustration, we find that the expected payoff to the seller under full disclosure equals $1/6$. The expected payoff under no disclosure is $1/8$. And the expected payoff under the optimal disclosure rule just described equals 0.1885. (All the relevant calculations, including details on the numerical exercise, are available in the Appendix.)

3.2.2 Example: Optimal Disclosure under Concave Demand Function

Again, suppose the object's value and it's profitability are independently distributed, each distributed according to the uniform distribution $U[0,1]$. Moreover, signal realizations are fully revealing, so that

$F_Y = U[0,1]$ and $F_{X|y} = U[0,1]$ for each $y \in [0,1]$. But now consider the following concave demand function: $p(x) = 2x - x^2$.

In this case, the thresholds defining the optimal disclosure rule should satisfy:

$$\bar{x} = x^{ND}, \text{ and}$$

$$\bar{y}(x) = y^{ND}\left[\frac{(2-2x^{ND})(x^{ND}-x)}{2x^{ND}-(x^{ND})^2-(2x-x^2)}\right] = y^{ND}\left[\frac{2(1-x^{ND})}{2-x^{ND}-x}\right].$$

Again, it must be that $x^{ND} = \mathbb{E}(x|\text{no disclosure})$, and $y^{ND} = \mathbb{E}(y|\text{no disclosure})$. Numerically, I find that the is a unique pair $(x^{ND}, y^{ND}) \in (0,1)^2$ satisfying these conditions, and therefore corresponding to the thresholds in the optimal disclosure rule, is

$$x^{ND} = 0.5397 \text{ and } y^{ND} = 0.5402.$$

For an illustration, I find that the expected payoff to the seller under full disclosure equals $1/3$. The expected payoff under no disclosure is $3/8$. And the expected payoff under the optimal disclosure rule just described equals 0.3907. (Again, all the relevant calculations, including details on the numerical exercise, are available in the Appendix.)

4 Mandated Transparency Policy

Consider a policy intervention that makes the motives of the seller transparent to the buyer. The policy is a transparency mandate imposed on the seller by a regulator, forcing them to always reveal their interest in the object's sale. Such transparency policies are the most common regulations used in advice markets; for example, regulators require financial advisors to disclose their commission structures (as mentioned in section 2.2), and platforms such as Instagram require influencers to tag "sponsored posts."

In the model, the policy imposes that when the buyer observes the seller's "advice" — either the disclosed signal realization or non-disclosure — they also perfectly observe how profitable the object's sale is to the seller. Naturally, anticipating that the buyer will see the profitability of the object, the seller optimally uses disclosure strategy which differs from the payoff maximizing disclosure rule without transparency. Indeed, with transparency, the seller's problem becomes continuum of separate disclosure problems, one for each profitability. This fact is stated in Theorem 2, along with a characterization of optimal disclosure under mandated transparency.

Theorem 2. *With mandated transparency, the seller chooses for each $y \in \mathcal{Y}$ a policy $d(\cdot, y) : \mathcal{X} \to [0,1]$ to maximize $P(y, d(\cdot, y))$, where*

$$P(y, d(\cdot, y)) = \int_{\mathcal{X}} \left[d(x,y)p(x) + (1 - d(x,y))p(x_y^{ND})\right] dF_{X|y}(x)$$

$$\textit{and } x_y^{ND} = \frac{\int_{\mathcal{X}}[1-d(x,y)]x dF_{X|y}(x)}{\int_{\mathcal{X}}[1-d(x,y)]dF_{X|y}(x)}. \tag{10}$$

Under mandated transparency, any optimal disclosure rule d^* almost everywhere satisfies

$$p(x) > p\left(x_y^{ND}\right) + p'\left(x_y^{ND}\right)\left(x - x_y^{ND}\right) \Rightarrow d^*(x,y) = 1, \tag{11}$$

$$\textit{and } p(x) < p\left(x_y^{ND}\right) + p'\left(x_y^{ND}\right)\left(x - x_y^{ND}\right) \Rightarrow d^*(x,y) = 0. \tag{12}$$

Under transparency, the seller cannot use strategic disclosure in order to steer sales across objects with different profitability levels. Rather, Theorem 2 shows that they separately maximize the probability that the object with each profitability level gets sold. The optimal disclosure rule may differ across objects with different profitabilities because $F_{X|y}$ may depend on y, which therefore implies the optimal value of x_y^{ND} — which determines the optimal disclosure rule, as per (11) and (12) — varies with y.

4.1 Curvature of p and the Effectiveness of Mandated Transparency

Proposition 2 below is a direct consequence of Theorem 2, and shows that mandating transparency of the seller's motives may increase or decrease the amount of evidence the seller voluntarily discloses to the buyer, and that this effect depends on the curvature of the buyer's demand function p.[12]

Proposition 2. *Let d and d' be optimal disclosure policies to the seller under mandated transparency and hidden motives, respectively.*

1. *If $p(\cdot)$ is strictly **convex**, mandated transparency improves evidence disclosure, that is, d has **more disclosure** than d'.*
2. *If $p(\cdot)$ is strictly **concave**, mandated transparency harms evidence disclosure, that is, d has **less disclosure** than d'.*

If the demand function is convex, then the mandated transparency policy increases the set of evidence pieces that the seller would voluntarily choose to disclose to the buyer. In fact, given that demand regime, any optimal disclosure policy under mandated transparency involves the voluntary disclosure of *all evidence* – so that the optimal d involves full disclosure, regardless of the object's profitability. To see this, note that for any value of x_y^{ND}, strict convexity of p implies that the condition on the right-hand side of (11) is satisfied strictly for any $x \neq x_y^{ND}$. An optimal disclosure rule must therefore satisfy $d^*(x,y) = 1$ almost everywhere.

[12] If the demand function is affine, a case that is not contemplated in Proposition 2, then under mandated transparency, the seller is indifferent between all disclosure rules, including full disclosure and no disclosure. The effect of mandated transparency on disclosure is therefore indeterminate.

Conversely, if the demand function is concave, then mandated transparency has the opposite effect on evidence disclosure: the optimal disclosure policy is for the seller to not disclose any of the realized evidence. Again, this can be seen directly from (11). Because p is strictly concave, then for any x_y^{ND}, the condition on the right-hand side of (11) fails for all $x \neq x_y^{ND}$, which implies that almost all signal realizations are optimally concealed from the buyer.

To evaluate whether transparency is good or bad in terms of the informativeness of the seller's advice to the buyer, it is not enough to consider the evidence that is voluntarily disclosed by the seller. Rather, we must also consider the information that is "involuntarily" conveyed to the buyer directly when they observe the object's profitability. Remember that objects with different profitabilities may have different value (and evidence) distributions, and therefore some information is conveyed directly through the mandated transparency of the seller's motives. To evaluate the overall impact of the policy on the informativeness of the seller's advice to the buyer, we consider its impact on the distribution of the buyer's beliefs induced by the policy and the seller's optimal disclosure policy.

Denote the transparency policy by $\tau \in \{0, 1\}$, with $\tau = 1$ indicating the mandated transparency environment, and $\tau = 0$ the hidden motives environment. Given a disclosure rule d and transparency policy τ, let $F^B(\cdot, d, \tau)$ be the distribution of posterior means observed by the buyer.[13] We say the pair (d, τ) is *more informative than* the pair (d', τ') if $F^B(\cdot, d, \tau)$ is more informative than $F^B(\cdot, d', \tau')$ in the Blackwell order. By evaluating the policy in terms of its implied Blackwell informativeness to the buyer, I am agnostic as to what the buyer's objective is, and therefore on what is the "surplus" that the policy-maker wishes to maximize: the buyer's value under any decision problem (for which the mean posterior is a sufficient statistic) is higher under the more Blackwell informative policy.

Corollary 3.

1. *If $p(\cdot)$ is strictly* ***convex****, the seller is more informative under mandated transparency.*
2. *If $p(\cdot)$ is strictly* ***concave****, and $\mathbb{E}[x|y] = \mathbb{E}[x]$ for every $y \in \mathcal{Y}$, then the seller is less informative under mandated transparency.*

Proposition 2 and its Corollary 3 show that the effectiveness of mandated transparency as a regulatory policy depends on the curvature of the demand function p. Returning to the micro-foundations provided in section 2.1, one possible interpretation of the curvature of the demand function is as a measure of the competitiveness in the market to which the seller belongs. Specifically, if there are sufficiently many "competitors" to the considered seller, then the demand function p is convex, and mandated transparency is an effective policy. Conversely, if competition is not sufficient, then p may be concave, implying

[13] For a given d and τ, $F^B(x, d, \tau)$ is equal to

$$\int_{\mathcal{Y}} \int_{[x_{min}, x)} d(\hat{x}, y) dF_{X|y}(\hat{x}) dF_Y(y) + \int_{\mathcal{Y}} \int_{\mathcal{X}} (1 - d(\hat{x}, y)) \mathbb{1}\{x_y^{ND} \leqslant x\} dF_{X|y}(\hat{x}) dF_Y(y),$$

where for every $y \in \mathcal{Y}$, $x_y^{ND} = x^{ND}$ as given in (1) if $\tau = 0$; and x_y^{ND} is as defined in (10) if $\tau = 1$.

that mandated transparency is not a good policy tool. An interpretation is that competition with outside sellers is a sufficient motivator for the seller to provide information to the buyer — this point is indeed made by Hwang, Kim and Boleslavsky (2023). But, in case there is lack of competition, allowing sellers to profitably steer buyers (by keeping their motives hidden) can provide the necessary incentives for the disclosure of information about the object's value.

4.2 Local Effects of Mandated Transparency

The results in section 4.1 concern extreme cases where p is strictly concave or strictly convex. "Local versions" of those results hold, which depend only on the local curvature of the demand function around the expected value x^{ND}. From (10), we know that under mandated transparency, a signal realization x for an object with profitability y is disclosed if

$$p(x) > p\left(x_y^{ND}\right) + p'\left(x_y^{ND}\right)\left(x - x_y^{ND}\right), \tag{13}$$

and x is not disclosed if the opposite inequality holds. Without mandated transparency, we know from (a rewriting of) the characterization in Theorem 1 that a signal realization x for an object with profitability y is disclosed if

$$p(x) > p\left(x^{ND}\right) + \frac{y^{ND}}{y} p'\left(x^{ND}\right)\left(x - x^{ND}\right), \tag{14}$$

where remember that y^{ND} is the expected profitability conditional on no disclosure under the optimal disclosure rule. Again, the signal realization x is not disclosed if the opposite inequality holds.

Suppose optimal disclosure rules with and without mandated transparency are such that, for some profitability level y, we have $x^{ND} = x_y^{ND}$. And further suppose that p is locally strictly convex around x^{ND}. In that case, condition (13) must hold for signal realizations x sufficiently close to x^{ND}, meaning that such "local signal realizations" are disclosed to the buyer. In comparison, consider the optimal policy when the seller has hidden motives. For high-profitability objects ($y > y^{ND}$) and "local bad news realizations" ($x \uparrow x^{ND}$), the opposite inequality to (14) must hold; so that such local bad news are concealed. Analogously, for low-profitability objects ($y < y^{ND}$), "local good news realizations" ($x \downarrow x^{ND}$) are concealed from the buyer. This argument implies that, if p is locally convex around x^{ND}, then mandated transparency implies a local increase in disclosure. This result is formally stated below in Proposition 3, along with an opposite result for the case when p is locally concave.

Proposition 3. *Suppose d and d_m are optimal disclosure rules with hidden motives and under mandated transparency, respectively. And suppose, for all $y \in \mathcal{Y}$, it holds that $x^{ND}(d) = x_y^{ND}(d_m) =: \tilde{x} \in int(\mathcal{X})$.*

1. *If p is locally strictly convex around $\tilde{x}$, there exist x' and x'', with $x' < \tilde{x} < x''$ such that $d(x, y) \leqslant d_m(x, y)$ for almost all (x, y) with $x \in (x', x'')$; and strictly so for an open subset of such (x, y).*

2. *If p is locally strictly concave around $\tilde{x}$, there exist x' and x'', with $x' < \tilde{x} < x''$ such that*

$d(x, y) \geqslant d_m(x, y)$ for almost all (x, y) with $x \in (x', x'')$; and strictly so for an open subset of such (x, y).

Also from conditions (13) and (14), we can see that, compared to the benchmark with hidden motives, mandated transparency implies a weak increase in the disclosure of bad news, and a weak decrease in the disclosure of good news, about objects with high profitability. To see, if the object's profitability is high ($y > y^{ND}$), and a signal realization is "bad news" ($x < x^{ND}$), then condition (14) holding implies that condition (13) holds as well. This means that if such bad news are disclosed under hidden motives, then they are disclosed as well under mandated transparency. Conversely, for "good news" ($x > x^{ND}$), condition (13) implies condition (14), so that their disclosure under mandated transparency implies that they are also disclosed under hidden motives. Proposition 4 states this result formally, as well as an analogous result for low profitability objects ($y < y^{ND}$).

Proposition 4. *Suppose d and d_m are optimal disclosure rules with hidden motives and under mandated transparency, respectively. And suppose, for all $y \in \mathcal{Y}$, it holds that $x^{ND}(d) = x_y^{ND}(d_m) =: \tilde{x}$; and let $\tilde{y} := y^{ND}(d)$. The following statement holds for almost all $(x, y) \in \mathcal{X} \times \mathcal{Y}$:*

$$d(x, y) \geqslant d_m(x, y) \Leftrightarrow (x - \tilde{x})(y - \tilde{y}) \geqslant 0.$$

Both Propositions 3 and 4 start from the assumption that optimal disclosure rules with and without mandated transparency lead to the same beliefs of no disclosure $x^{ND} = x_y^{ND}$ (for all profitability levels $y \in \mathcal{Y}$). If the distribution of signal realizations $F_{X|y}$ is independent from y, then we know from Theorem 2 that in the optimal disclosure rule under mandated disclosure, we have that x_y^{ND} is also independent of y. Moreover, as the distribution of profitabilities becomes more centered around its expected value, it must be that the value x^{ND} in the optimal rule under hidden motives approaches such value x_y^{ND}. Indeed, in the limit as the distribution F_Y becomes the degenerate distribution, it must be that these two values coincide.

5 Alternative Communication Protocols

My model makes two main assumptions regarding the communication protocol. First, the seller can *commit* to a rule to disclose signal realizations, prior to the object's profitability or the "evidence" about its value being drawn. Second, the seller can use only disclosure policies, either revealing or not revealing a realized piece of information about the object's value, rather than committing to more general signal structures. In this section, I consider a variations of the model, which drops the commitment from the communication protocol. Under that new protocol, I investigate whether mandating transparency about the seller's motives improves the informativeness of their advice to the buyer. Next, I briefly comment on a variation of the model in which the seller can commit to more general signal structures.

5.1 No Commitment Disclosure Protocol

A direct reading of the results in section 4 is that they delineate conditions under which a policy maker should or should not institute a transparency policy: such decision should be made based on the curvature of the buyer's demand for the object. But beyond this normative implication, the results also refine our theoretical understanding of regulation in advice markets. Specifically, I highlight that in information design models *with commitment*, the 'alignment between sender and receiver preferences' and the 'opaqueness of the sender's motives' are distinct objects; and it is not necessarily true that regulations that reduce the latter would also make the sender's interests more aligned with those of the receiver. And even in contexts where transparency does improve the alignment between the advisor and advisee's interests — such as when the demand function is convex, in the model — it does so because, as a byproduct the transparency of their motives, the sender's effective objective function becomes "more convex," therefore inducing them to optimally disclose more evidence. This is in contrast with what happens in a disclosure environment without commitment — discussed below — in which the transparency of the seller's motives induces the unravelling of uninformative equilibria.

To see this contrast, I now introduce a version of the model where there is no commitment in the communication protocol. In this section, we allow the support of the object's profitability to include negative profitability values, so we may have $y_{min} < 0 < y_{max}$,[14] in which case the buyer is unsure whether the profitability is such that the seller wishes to maximize the object's probability of sale or to minimize it. Consider the following disclosure protocol with no commitment: First, the seller observes the object's profitability and a signal realization. After that, the seller chooses whether to disclose the signal realization (a piece of *evidence* about the object's value) to the buyer. The buyer observes the disclosed evidence – or the fact that no evidence was disclosed – and forms a posterior mean about the object's value, taking into account the seller's equilibrium strategy. As before, the buyer does not observe the object's profitability.The equilibrium notion is Perfect Bayesian Equilibrium.

Proposition 5. *For any demand function* p*, an equilibrium exists, and any equilibrium disclosure strategy* d^* *has a threshold structure:* $d^*(x, y) \in \{0, 1\}$ *and*[15]

$$d^*(x, y) = 1 \Leftrightarrow (x - \bar{x})y \geqslant 0,$$

for some $\bar{x} \in \mathcal{X}$*, satisfying* $\bar{x} = \mathbb{E}[x|(x - \bar{x})y < 0]$ *if* $\{(x - \bar{x})y < 0\} \neq \varnothing$.

Suppose the object's profitability is always positive ($y_{min} \geqslant 0$), and conjecture a threshold equilibrium with some evidence concealment – that is, suppose $\bar{x} \in (x_{min}, x_{max})$. Then, because y is always greater than 0, it must be that $\bar{x} > \mathbb{E}[x|(x - \bar{x})y < 0]$, and thus the equilibrium condition in Proposition 5 is not satisfied. Such a conjectured equilibrium would unravel: the buyer's posterior upon observing non-disclosure would be $\mathbb{E}[x|(x - \bar{x})y < 0]$, which is strictly smaller than $\bar{x}$. Consequently, when the

[14] Note that the characterization of optimal disclosure rules under the commitment protocol, given in Theorem 1, also applies when $y_{min} < 0$.

[15] Assuming that, when indifferent, the seller discloses the signal realization.

seller draws a signal realization just under $\bar{x}$, they strictly prefer to reveal it to the buyer, which is inconsistent with the initially conjectured equilibrium. An analogous unravelling argument applies if the object's profitability is always negative ($y_{max} \leqslant 0$). However, when profitability can be both positive or negative, so that $0 \in (y_{min}, y_{max})$, then any equilibrium involves partial disclosure. There is some interior threshold $\bar{x}$ such that the seller discloses (only and) all "good news" about the object ($x \geqslant \bar{x}$) when profitability is positive, and (only and) all "bad news" ($x \leqslant \bar{x}$) about the object when profitability is negative. Partially uninformative equilibria do not unravel, because, upon observing non-disclosure, the buyer does not know whether the seller has "good news" but negative profitability or "bad news" but positive profitability.[16]

Note also that equilibrium disclosure strategies as described in the proposition are independent of the shape of the demand function p – in contrast to the model with commitment, where the shape of optimal disclosure rules depends on the curvature of p. This happens because the seller makes their disclosure decision only after seeing the signal realization; and at that point their best response is guided solely from comparing the given realization to the buyer's belief of no disclosure. If the profitability of the object is positive, then, because the demand function is strictly increasing, the best response is to disclose good news (better than the belief of no disclosure), and conceal otherwise. If instead the profitability of the object is negative, the best response is to disclose bad news, and conceal otherwise. This implies that the "threshold profitability" is always 0, which divides positive profitability objects from negative profitability objects.

Now return to the question of whether mandated transparency incentivizes the seller to disclose information about the object's value to the buyer. Proposition 6 shows that, in contrast with the benchmark with commitment, in this case mandated transparency induces the seller to reveal all their evidence about the object's value; and this result holds independently of the curvature of the buyer's demand function.

Proposition 6. *Under a mandated transparency policy that reveals the object's profitability to the buyer, full disclosure is the unique equilibrium of the disclosure game with no commitment.*

5.2 Unconstrained Signaling Technology

The usual assumption in information design models is that a sender (the seller, in this case) commits a signal a map from states of the world (the signal realizations about the object's value) into distributions of messages to inform a receiver (the buyer) about the state. The sender's choice of such a map is unrestricted. Contrastingly, in this paper, I assumed that the sender is restricted to a class of "signaling strategies:" the class of simple disclosure rules, in which the sender's message either fully conveys the information in a piece of evidence, or is "silent." Such silence is a message in itself, which conveys to the receiver that the state of the world is "one of the value-profitability pairs that would lead the sender to stay silent."

[16] The existence of partly-uninformative equilibria as described in Proposition 5 is in line with Seidmann and Winter's (1997) observation that in disclosure environments where the sender's preferred action depends on their type, there may exist equilibria in which their type is not always revealed in equilibrium.

In the *constrained information design* problem I study, Theorem 1 provides quite a complete characterization of the seller's optimal messaging strategy. In comparison, a characterization of the sender's optimal signal in the equivalent unconstrained design problem is elusive – Rayo and Segal (2010) provide a partial characterization of the optimal signal when the demand function p is affine. Regardless, some of the results in this paper also hold in the "unconstrained design version" of the problem. For example, Proposition 2, and its Corollary 3 would still be true, so that transparency may be detrimental to the seller's incentives to relay information about the object's value to the buyer, depending on the curvature of the buyer's demand function.

6 Conclusion

This paper models an environment in which an advisee seek information provided by an advisor with hidden motives. The advisee is a buyer who is informed by a seller about the value of an object that is on sale. The buyer understands that the seller is biased towards pushing the sale of the product, but does not know the extent of the seller's interest (the degree to which the object's sale is profitable to the seller). This environment describes many possible applied contexts, including the relation between an investor and their financial advisor, or that between a "follower" and their preferred social media influencer.

From a theoretical perspective, the first contribution of the paper is characterizing optimal disclosure rules in an environment where the seller with "hidden motives" — as in Rayo and Segal (2010) — can commit to a rule to disclose realizations of a signal about the object's value to the buyer. Theorem 1 shows that such rules generally have a threshold structure: there exists a *threshold value* and a *threshold profitability function* that define four "quadrants" in the value $\times$ profitability space. The seller optimally discloses realizations in the first and third quadrants — in which seller and buyer have aligned interests — and conceals realizations in the second and fourth quadrant, which correspond to misaligned interests realizations. The subsequent analysis shows how features of the threshold profitability function are determined by the curvature of the buyer's demand function.

I contribute also to a regulatory debate, regarding whether policies should be instituted that require advisors to reveal their interests to their advisees. Such "mandated transparency" policies are the most common regulations proposed to address the mis-alignment between advisor and advisee's interests; for example, the SEC mandates that financial advisors disclose commissions received from financial product providers, and the FTC and UK CMA advise social media influencers to clarify their relationships with brands to their followers. Proposition 2 argues that the effectiveness of a transparency policy can depend on the curvature of the buyer's demand function, highlighting that there is no "one size fits all" policy to optimally regulate advice markets. Rather, it is important to fit the regulation to specific features of each market. Using the interpretation introduced in Section 2.1, which relates the curvature of the buyer's demand function to the degree of competition in the relevant market, we learn that mandating transparency is an effective policy in very competitive markets, but may not be so in markets where the buyer does not have many alternatives to the object being offered by the seller.

Finally, I study a variation of the benchmark communication protocol of "disclosure with commitment." Specifically, I show that if the seller communicates through disclosure of verifiable information *without commitment*, then mandated transparency is an effective policy, inducing the seller to be perfectly informative regardless of the shape of the buyer's demand function. To the degree that commitment is a stand-in assumption representing the degree of recurrence in the sender-receiver relationship, my analysis shows that the importance of reputation-building between advisor and advisee also impacts the optimal regulation of advice markets.

References

Anagol, S., S. Cole, and S. Sarkar. (2017) "Understanding the Advice of Commissions Motivated Agents: Evidence from the Indian Life Insurance Market." *Review of Economics and Statistics*, **99**, 1-15.

Ball, I., and X. Gao. (2024) "Benefiting from Bias: Delegating to Encourage Information Acquisition." *Journal of Economic Theory*.

Bar-Isaac, H., and J. Shapiro. (2013) "Ratings Quality over the Business Cycle." *Journal of Financial Economics*, **108**, 62-78.

Best, J., and D. Quigley. (2023) "Persuasion for the Long Run." *Journal of Political Economy*, forthcoming.

Bolton, P., X. Freixas, and J. Shapiro. (2012) "The Credit Ratings Game." *The Journal of Finance*, **67**, 85-111.

Chalmers, J. and J. Reuter. (2020) "Is Conflicted Investment Advice Better than no Advice?." *Journal of Financial Economics*, forthcoming.

Che, Y.K., and N. Kartik (2009) "Opinions as Incentives." *Journal of Political Economy*, **117**, 815-860.

Crawford, V. and J. Sobel. (1982) "Strategic Information Transmission." *Econometrica*: 1431-1451.

Deb, R., M. Pai, and M. Said (2023). "Indirect Persuasion," *working paper*.

DeMarzo, P. M., I. Kremer and A. Skrzypacz (2019) "Test Design and Minimum Standards." *American Economic Review*, **109**: 2173-2207.

Doval, L., and V. Skreta. (2024) "Constrained Information Design," *Mathematics of Operations Research* **49**: 78-106..

Dranove, D. and G. Jin. (2010) "Quality Disclosure and Certification: Theory and Practice." *Journal of Economic Literature*, **48**: 935-63.

Dye, R. A. (1985) “Disclosure of Nonproprietary Information.” *Journal of Accounting Research*, **23**: 123-145.

Eckardt, M., and S. Rathke-Doppner. (2010) “The Quality of Insurance Intermediary Services – Empirical Evidence for Germany.” *Journal of Risk and Insurance*, **77**, 667-701.

Ershov, D., Y. He, and S. Seiler. (2023) “How Much Influencer Marketing is Undisclosed? Evidence from Twitter,” *working paper*.

Ershov, D. and M. Mitchell. (2023) “The Effects of Advertising Disclosure Regulations on Social Media: Evidence From Instagram.” *RAND Journal of Economics*, forthcoming.

Grossman, S. (1981) “The Informational Role of Warranties and Private Disclosure about Product Quality.” *Journal of Law & Economics*, **24**: 461-484.

Hwang, I., K. Kim, and R. Boleslavsky (2023). “Competitive Advertising and Pricing,” *working paper*.

Inderst, R., and M. Ottaviani. (2012.1) “Competition through Commissions and Kickbacks.” *American Economic Review*, **102**, 780-809.

Inderst, R., and M. Ottaviani. (2012.2) “Financial Advice”. *Journal of Economic Literature*, **50**, 494-512.

Inderst, R., and M. Ottaviani. (2012.3) “How (not) to Pay for Advice: A Framework for Consumer Financial Protection.” *Journal of Financial Economics*, **105**, 393-411.

Jung, W.-O. and Y. K. Kwon (1988) “Disclosure When the Market Is Unsure of Information Endowment of Managers.” *Journal of Accounting Research*, **26**: 146-153.

Kamenica, E., and M. Gentzkow. (2011) “Bayesian Persuasion.” *American Economic Review*, **101**, 2590-2615.

Kartik, N., F. X. Lee, and W. Suen. (2017) “Investment in Concealable Information by Biased Experts.” *The RAND Journal of Economics*, **48**: 24-43.

Li, M. and K. Madarász. (2008) “When Mandatory Disclosure Hurts: Expert Advice and Conflicting Interests.” *Journal of Economic Theory* **139**: 47-74.

Libgober, J. (2022) ”False Positives and Transparency.” *American Economic Journal: Microeconomics* **14**: 478-505.

Lipnowski, E. and D. Ravid. (2020) “Cheap Talk with Transparent Motives.” *Econometrica*, **88**: 1631-1660.

Lin, X. and C. Liu. (2023) “Credible Persuasion.” *Journal of Political Economy*, forthcoming.

Mathevet, L., D. Pearce, and E. Stacchetti. (2024) "Reputation and Information Design," *working paper*.

Matthews, Steven, and Andrew Postlewaite. (1985) "Quality Testing and Disclosure." *RAND Journal of Economics*: 328-340.

Mensch, J. (2021) "Monotone Persuasion." *Games and Economic Behavior*, **130**: 521-542.

Milgrom, P. (1981) "Good News and Bad News: Representation Theorems and Applications." *The Bell Journal of Economics*, **12**: 380-391.

Milgrom, P. (2008) "What the Seller Won't Tell You: Persuasion and Disclosure in Markets." *Journal of Economic Perspectives*, **22**: 115?131.

Morgan, J. and P. Stocken. (2003) "An Analysis of Stock Recommendations." *RAND Journal of Economics*: 183-203.

Morris, S. (2001) "Political Correctness." *Journal of Political Economy* **109**: 231-265.

Onuchic, P., and D. Ray. (2022) "Conveying Value via Categories," *Theoretical Economics*, forthcoming.

Opp, C., M. Opp, and M. Harris. (2013) "Rating Agencies in the Face of Regulation." *Journal of Financial Economics*, **108**, 46-61.

Rayo, L., and I. Segal (2010) "Optimal Information Disclosure." *Journal of Political Economy*, **118**, 949–987.

Seidmann, D., and E. Winter. (1997) "Strategic Information Transmission with Verifiable Messages." *Econometrica*: 163-169.

Shishkin, D. (2023) "Evidence Acquisition and Voluntary Disclosure." *working paper*.

Skreta, V., and L. Veldkamp. (2009) "Ratings Shopping and Asset Complexity: A Theory of Ratings Inflation." *Journal of Monetary Economics*, **56**, 678-695.

Sobel, J. (1985) "A Theory of Credibility." *Review of Economic Studies* **52**: 557-573.

Szalay, D. (2005) "The Economics of Clear Advice and Extreme Options." *The Review of Economic Studies*, **72**: 1173-1198.

A Proofs

Statements and Proofs of Claims 1 and 2

This section completes the proof in the main text, showing that any disclosure rule that does not satisfy the threshold structure described in (4) can be improved upon by a rule that does satisfy (4). To that end, consider d and $\hat{d}$ as in (6) and (7). I prove the following two claims used in the main text.

Claim 1. *d and $\hat{d}$ produce the same overall probability of sale.*

Proof.

$$
\begin{aligned}
&\mathbb{E}[P(y,\hat{d})] - \mathbb{E}[P(y,d)] = \\
&\qquad = \int_{\mathcal{Y}} \int_{\mathcal{X}} \left[p(x) - p(x^{ND})\right] \left[\hat{d}(x,y) - d(x,y)\right] dF_{X|y}(x) dF_Y(y) \\
&\qquad = \int_{\mathcal{X}} \left[p(x) - p(x^{ND})\right] \int_{\mathcal{Y}} \left[\hat{d}(x,y) - d(x,y)\right] dF_{Y|x}(y) dF_X(x) = 0
\end{aligned}
$$

where $F_{Y|x}$ is the profitability distribution conditional on a signal realization x and F_X is the marginal distribution of signal realizations. The first equality uses the definition of $P(y,d)$ and the third is due to d and $\hat{d}$ disclosing each realization with the same probability, as in (8). ■

Claim 2. *$\hat{d}$ induces a larger covariance between sales and profitability than d.*

Proof.

$$
\begin{aligned}
&\text{Cov}\left[y, P(y,\hat{d})\right] - \text{Cov}\left[y, P(y,\hat{d})\right] = \mathbb{E}\left[\left(P(y,\hat{d}) - P(y,d)\right)(y - \mathbb{E}(y))\right] \\
&= \int_{\mathcal{Y}} \int_{\mathcal{X}} \left[p(x) - p(x^{ND})\right] \left[\hat{d}(x,y) - d(x,y)\right] [w - \mathbb{E}(y)]\, dF_{X|y}(x) dF_Y(y) \\
&= \int_{\mathcal{X}} \left[p(x) - p(x^{ND})\right] \int_{\mathcal{Y}} \left[\hat{d}(x,y) - d(x,y)\right] [y - \mathbb{E}(y)]\, dF_{Y|x}(y) dF_X(x)
\end{aligned} \tag{15}
$$

By the definition of $\hat{d}$, for $x < x^{ND}$, $\hat{d}(x,y) - d(x,y) \geqslant 0$ when $y < \hat{y}(x)$ and $\hat{d}(x,y) - d(x,y) \leqslant 0$ when $y > \hat{y}(x)$ – but, as given by (8), the expected difference $\hat{d}(x,y) - d(x,y)$ is 0. This, along with the fact that $y - \mathbb{E}(y)$ is increasing in y, implies that

$$
\int_{\mathcal{Y}} \left[\hat{d}(x,y) - d(x,y)\right] [y - \mathbb{E}(y)]\, dF_{Y|x}(y) \leqslant 0,
$$

when $x \leqslant x^{ND}$. Analogously, we can show that

$$
\int_{\mathcal{Y}} \left[\hat{d}(x,y) - d(x,y)\right] [w - \mathbb{E}(y)]\, dF_{Y|x}(y) \geqslant 0,
$$

when $x > x^{ND}$. Moreover, these inequalities are strict for a positive measure of signal realizations. These observations, along with the fact that p is strictly increasing, deliver that the expression in (15) is strictly positive. ■

Proof of Theorem 1

Step 1. Suppose $\hat{d}$ is a disclosure rule such that no disclosure happens with positive probability — that is, $\int_{\mathcal{Y}} \int_{\mathcal{X}} \left[1 - \hat{d}(x,y)\right] dF_{X|y}(x) dF_Y(y) > 0$ — and suppose $\hat{d}$ does not satisfy the characterization in

Theorem 1. Then d can be strictly improved.

For any disclosure rule d such that no disclosure happens with positive probability, the seller's value is given by

$$\int_y yP(y,d)dF_Y(y) = \int_y y \int_{\mathcal{X}} \left[d(x,y)p(x) + (1-d(x,y))p(x^{ND})\right] dF_{X|y}(x)dF_Y(y),$$

$$\text{where } x^{ND} = \frac{\int_{\mathcal{Y}}\int_{\mathcal{X}} x\,(1-d(x,y))\,dF_{X|y}(x)dF_Y(y)}{\int_{\mathcal{Y}}\int_{\mathcal{X}} (1-d(x,y))\,dF_{X|y}(x)dF_Y(y)}.$$

For $y \in \mathcal{Y}$ and $x \in \mathcal{X}$, we can take a derivative of the sender's value with respect to $d(x,y)$, to get

$$\begin{aligned}\frac{\partial \Pi}{\partial d(x,y)} =& y\left(p(x) - p(x^{ND})\right) dF_{X|y}(x)dF_Y(y) \\ &+ \left(\int_{\mathcal{Y}}\int_{\mathcal{X}} \tilde{y}\,[1-d(\tilde{x},\tilde{y})]\,dF_{X|\tilde{y}}(\tilde{x})dF_Y(\tilde{y})\right) p'(x^{ND})\frac{\partial x^{ND}}{\partial d(x,y)}\end{aligned}$$

Now from the definition of x^{ND}, we get

$$\frac{\partial x^{ND}}{\partial d(x,y)} = \frac{\int_{\mathcal{Y}}\int_{\mathcal{X}}(\tilde{x}-x)(1-d(\tilde{x},\tilde{y}))dF_{X|\tilde{y}}(\tilde{x})dF_Y(\tilde{y})}{\left(\int_{\mathcal{Y}}\int_{\mathcal{X}}(1-d(\tilde{x},\tilde{y}))dF_{X|\tilde{y}}(\tilde{x})dF_Y(\tilde{y})\right)^2} dF_{X|y}(x)dF_Y(y)$$

Substituting this into the previous equation, we have

$$\frac{\partial \Pi}{\partial d(x,y)} = \left[y\left(p(x)-p(x^{ND})\right) - y^{ND}p'(x^{ND})(x-x^{ND})\right] dF_{X|y}(x)dF_Y(y), \tag{16}$$

where y^{ND} is the average object profitability given non-disclosure. It is easy to check that, if $x < x^{ND}$,

$$\frac{\partial \Pi}{\partial d(x,y)} \begin{cases} > 0, \text{ if } y < y^{ND}\left[\frac{p'(x^{ND})(x^{ND}-x)}{p(x^{ND})-p(x)}\right] \\ < 0, \text{ if } y > y^{ND}\left[\frac{p'(x^{ND})(x^{ND}-x)}{p(x^{ND})-p(x)}\right] \end{cases}$$

Conversely, if $x > x^{ND}$,

$$\frac{\partial \Pi}{\partial d(x,y)} \begin{cases} > 0, \text{ if } y > y^{ND}\left[\frac{p'(x^{ND})(x^{ND}-x)}{p(x^{ND})-p(x)}\right] \\ < 0, \text{ if } y < y^{ND}\left[\frac{p'(x^{ND})(x^{ND}-x)}{p(x^{ND})-p(x)}\right] \end{cases}$$

Now take disclosure $\hat{d}$, which does not satisfy the characterization in Theorem 1. That is, $\hat{d}$ does not have a threshold structure as in (4) with $\bar{x} = x^{ND}$ and profitability threshold satisfying

$$\bar{y}(x) = y^{ND}\left[\frac{p'(x^{ND})(x^{ND}-x)}{p(x^{ND})-p(x)}\right].$$

Then it must be that either $\hat{d}(x, y) \neq 0$ when (16) is negative or $\hat{d}(x, y) \neq 1$ when (16) is positive. In each case, $\hat{d}$ can be strictly improved.

Step 2. Suppose instead that $\hat{d}$ is the "full disclosure" rule — that is, $d(x, y) = 1$ almost everywhere. Then there are two possibilities. First, it may be that full disclosure is a rule that satisfies (4), (5), and $\bar{x} = x^{ND}$ — for $x^{ND} = \inf(\mathcal{X})$ and $y^{ND} = \inf(\mathcal{Y})$ or $x^{ND} = \sup(\mathcal{X})$ and $y^{ND} = \sup(\mathcal{Y})$. In that case, full disclosure satisfies the necessary conditions for optimality, as given by the Theorem.

If instead full disclosure is not a rule that satisfies the necessary conditions given in the Theorem, it must be that the disclosure rule implied by (4), (5), and $\bar{x} = x^{ND}$, with $x^{ND} = \inf(\mathcal{X})$ and $y^{ND} = \inf(\mathcal{Y})$, is such that no disclosure happens with positive probability. For this to be the case, there must exist $x \in \mathcal{X}$ such that

$$\frac{p'\left(\inf(\mathcal{X})\right)\left(x - \inf(\mathcal{X})\right)}{p(x) - p\left(\inf(\mathcal{X})\right)} > 1.$$

Similarly, it must be that the disclosure rule implied by (4), (5), and $\bar{x} = x^{ND}$, with $x^{ND} = \sup(\mathcal{X})$ and $y^{ND} = \sup(\mathcal{Y})$, is such that no disclosure happens with positive probability. For this to be the case, there must exist $x' \in \mathcal{X}$ such that

$$\frac{p'\left(\sup(\mathcal{X})\right)\left(x' - \sup(\mathcal{X})\right)}{p(x') - p\left(\sup(\mathcal{X})\right)} < 1.$$

When such x and x' exist, we know from Proposition 1 and Lemma 2 that the optimal disclosure rule must be interior, and so that no disclosure happens with positive probability. Therefore, $\hat{d}$, the "full disclosure" rule, is not optimal. ∎

Proof of Lemma 1

Suppose system (9) has a solution with $\hat{x} \in int(\mathcal{X})$ and $\hat{y} \in int(\mathcal{Y})$, and let $d_{\hat{x},\hat{y}}$ be the corresponding disclosure rule defined by such $\hat{x}$ and $\hat{y}$.

Step 1. Showing that $d_{\hat{x},\hat{y}}$ yields strictly higher value to the seller than full disclosure.

For each $\alpha \in [0, 1]$, define the following alternative disclosure rule:

$$1 - d_\alpha(x, y) = \begin{cases} 1 - d_{\hat{x},\hat{y}}(x, y), \text{ if } d_{\hat{x},\hat{y}}(x, y) = 1 \\ \alpha\left(1 - d_{\hat{x},\hat{y}}(x, y)\right), \text{ otherwise} \end{cases}$$

Note that, because for all $\alpha > 0$, the probability of no disclosure implied by d_α is proportional to that implied by $d_{\hat{x},\hat{y}}$, it must be that these rules imply the same x^{ND} and y^{ND}. Therefore, because $d_{\hat{x},\hat{y}}$ is such that $\hat{x}$ and $\hat{y}$ solve the system (9), we know that as α decreases, there is an increase in disclosure for realizations (x, y) such that $\partial\Pi/\partial d(x, y)$, as given by (16), is strictly negative. Therefore, the value of d_α to the seller is strictly increasing in α. Moreover, full disclosure corresponds to the case of $\alpha = 0$, which is therefore dominated by $d_{\hat{x},\hat{y}}$.

Step 2. Showing that any "corner disclosure rule" $d_{\tilde{x},\tilde{y}}$ defined by $\tilde{x} \in \{\inf(\mathcal{X}), \sup(\mathcal{X})\}$ or $\tilde{y} \in \{\inf(\mathcal{Y}), \sup(\mathcal{Y})\}$ such that no disclosure happens with positive probability cannot be an optimal disclosure rule.

For any disclosure rule such that no disclosure happens with positive probability, it must be that $\mathbb{E}(x|\text{ no disclosure}) \in int(\mathcal{X})$ and $\mathbb{E}(y|\text{ no disclosure}) \in int(\mathcal{Y})$ — because F has full support over $\mathcal{X} \times \mathcal{Y}$ and no mass points. Therefore, if $\tilde{x} \in \{\inf(\mathcal{X}), \sup(\mathcal{X})\}$ or $\tilde{y} \in \{\inf(\mathcal{Y}), \sup(\mathcal{Y})\}$, the system (9) is not satisfied, and by Theorem 1 disclosure rule $d_{\tilde{x},\tilde{y}}$ can be strictly improved.

■

Proof of Proposition 1

For each $\hat{x} \in \mathcal{X}$ and $\hat{y} \in \mathcal{Y}$, define $d_{\hat{x},\hat{y}}$ by

$$(x-\hat{x})(y-\bar{y}(x)) \geqslant 0 \Rightarrow d_{\hat{x},\hat{y}}(x,y)=1, \quad (x-\bar{x})(y-\bar{y}(x)) < 0 \Rightarrow d_{\hat{x},\hat{y}}(x,y)=0,$$

$$\text{where } \bar{y}(x) = \hat{y}\left[\frac{p'(\hat{x})(\hat{x}-x)}{p(\hat{x})-p(x)}\right].$$

Proof of Statement 1. Suppose there exist $x, x' \in \mathcal{X}$ such that

$$\frac{p'\left(\inf(\mathcal{X})\right)\left(x-\inf(\mathcal{X})\right)}{p(x)-p\left(\inf(\mathcal{X})\right)} > 1 \text{ and } \frac{p'\left(\sup(\mathcal{X})\right)\left(x'-\sup(\mathcal{X})\right)}{p(x')-p\left(\sup(\mathcal{X})\right)} < 1. \tag{17}$$

Note that, because (17) holds and p is continuously differentiable, for all $\hat{x} \in \mathcal{X}$ and $\hat{y} \in \mathcal{Y}$, $d_{\hat{x},\hat{y}}(x,y) = 0$ for some non-empty open subset of $\mathcal{X} \times \mathcal{Y}$. Therefore, $\mathbb{E}\left[x|\text{no disclosure region implied by } d_{\hat{x},\hat{y}}\right]$ and $\mathbb{E}\left[y|\text{no disclosure region implied by } d_{\hat{x},\hat{y}}\right]$ are well defined. Consider the following two continuous functions

$$\Gamma_1(\hat{x},\hat{y}) = \mathbb{E}\left[x|\text{no disclosure region implied by } d_{\hat{x},\hat{y}}\right] - \hat{x}$$

$$\text{and } \Gamma_2(\hat{x},\hat{y}) = \mathbb{E}\left[y|\text{no disclosure region implied by } d_{\hat{x},\hat{y}}\right] - \hat{y}.$$

For a disclosure rule such that no disclosure happens with positive probability, $\mathbb{E}(x|\text{ no disclosure}) \in int(\mathcal{X})$ and $\mathbb{E}(y|\text{ no disclosure}) \in int(\mathcal{Y})$ — because F has full support over $\mathcal{X} \times \mathcal{Y}$ and no mass points. Consequently, for every $\hat{y}$, $\Gamma_1\left(\inf(\mathcal{X}),\hat{y}\right) > 0$ and $\Gamma_1\left(\sup(\mathcal{X}),\hat{y}\right) < 0$. Similarly, for every $\hat{x}$, $\Gamma_2\left(\hat{x},\inf(\mathcal{Y})\right) > 0$ and $\Gamma_1\left(\hat{x},\sup(\mathcal{Y})\right) < 0$. Therefore, by the Pointcaré-Miranda Theorem, there exist $\hat{x} \in int(\mathcal{X})$ and $\hat{y} \in int(\mathcal{Y})$ such that both $\Gamma_1(\hat{x},\hat{y})$ and $\Gamma_2(\hat{x},\hat{y})$ are simultaneously equal to 0; which are solutions to system (9). ■

Proof of Statement 2. Suppose p is affine. The proof uses the following claim.

Claim 3. *Let $x_\epsilon = \inf(\mathcal{X}) + \epsilon$ and $y_\epsilon = \inf(\mathcal{Y}) + \epsilon$. Then*

$$\lim_{\epsilon \to 0} \mathbb{E}\left[x | \textit{no disclosure region implied by } d_{x_\epsilon, y_\epsilon}\right] > \inf(\mathcal{X}),$$

$$\textit{and } \lim_{\epsilon \to 0} \mathbb{E}\left[y | \textit{no disclosure region implied by } d_{x_\epsilon, y_\epsilon}\right] > \inf(\mathcal{Y}).$$

Proof. $\mathbb{E}\left[x | \text{no disclosure region implied by } d_{x_\epsilon, y_\epsilon}\right]$ is given by

$$\inf(\mathcal{X}) + \left[\int_{\inf(\mathcal{Y})}^{y_\epsilon} \int_{\epsilon}^{\Delta} x f_{X|y}(\inf(\mathcal{X}) + x) f_Y(y) dx dy + \int_{y_\epsilon}^{\sup(\mathcal{Y})} \int_0^{\epsilon} x f_{X|y}(\inf(\mathcal{X}) + x) f_Y(y) dx dy\right]$$

$$\times \left[\int_{\inf(\mathcal{Y})}^{y_\epsilon} \int_{\epsilon}^{\Delta} f_{X|y}(\inf(\mathcal{X}) + x) f_Y(y) dx dy + \int_{y_\epsilon}^{\sup(\mathcal{Y})} \int_0^{\epsilon} f_{X|y}(\inf(\mathcal{X}) + x) f_Y(y) dx dy\right]^{-1},$$

where I use the notation $\Delta = \sup(\mathcal{X}) - \inf(\mathcal{X})$. To assess the limit of this object as $\epsilon \to 0$, I use L'Hospital's rule. The derivative of the numerator is

$$\frac{\partial \text{NUM}}{\partial \epsilon} = \int_{\epsilon}^{\Delta} x f_{X|y_\epsilon}(\inf(\mathcal{X}) + x) f_Y(y_\epsilon) dx - \int_0^{\epsilon} x f_{X|y_\epsilon}(\inf(\mathcal{X}) + x) f_Y(y_\epsilon) dx$$
$$- \int_{\inf(\mathcal{Y})}^{y_\epsilon} \epsilon f_{X|y}(\inf(\mathcal{X}) + \epsilon) f_Y(y) dy + \int_{y_\epsilon}^{\sup(\mathcal{Y})} \epsilon f_{X|y}(\inf(\mathcal{X}) + \epsilon) f_Y(y) dy.$$

The derivative of the denominator is

$$\frac{\partial \text{DEN}}{\partial \epsilon} = \int_{\epsilon}^{\Delta} f_{X|y_\epsilon}(\inf(\mathcal{X}) + x) f_Y(y_\epsilon) dx - \int_0^{\epsilon} f_{X|y_\epsilon}(\inf(\mathcal{X}) + x) f_Y(y_\epsilon) dx$$
$$- \int_{\inf(\mathcal{Y})}^{y_\epsilon} f_{X|y}(\inf(\mathcal{X}) + \epsilon) f_Y(y) dy + \int_{y_\epsilon}^{\sup(\mathcal{Y})} f_{X|y}(\inf(\mathcal{X}) + \epsilon) f_Y(y) dy.$$

Therefore we have

$$\lim_{\epsilon \to 0} \mathbb{E}\left[x | \text{no disclosure region implied by } d_{x_\epsilon, y_\epsilon}\right] = \inf(\mathcal{X}) + \frac{\lim_{\epsilon \to 0} \frac{\partial \text{NUM}}{\partial \epsilon}}{\lim_{\epsilon \to 0} \frac{\partial \text{DEN}}{\partial \epsilon}} =$$

$$= \inf(\mathcal{X}) + \frac{\int_0^{\Delta} x f_{X|\inf(\mathcal{Y})}(\inf(\mathcal{X}) + x) f_Y(\inf(\mathcal{Y})) dx}{\int_0^{\Delta} f_{X|\inf(\mathcal{Y})}(\inf(\mathcal{X}) + x) f_Y(\inf(\mathcal{Y})) dx + \int_{\inf(\mathcal{Y})}^{\sup(\mathcal{Y})} f_{X|y}(\inf(\mathcal{X})) f_Y(y) dy} > \inf(\mathcal{X}),$$

where the inequality is due to f_Y and $f_{X|y}$, for each y, being strictly positive densities. The second part of the second observation — that $\lim_{\epsilon \to 0} \mathbb{E}\left[y | \text{no disclosure region implied by } d_{x_\epsilon, y_\epsilon}\right] > \inf(\mathcal{Y})$ — can be shown analogously. ∎

By Claim 3, we know that there exists some $\delta > 0$ such that $\Gamma_1(\inf(\mathcal{X}) + \delta, \inf(\mathcal{Y}) + \delta) > 0$ and

$\Gamma_2(\inf(\mathcal{X})+\delta, \inf(\mathcal{Y})+\delta) > 0$. Moreover, we have that for any $\hat{y} > \inf(\mathcal{Y})+\delta$, it also holds that

$$\mathbb{E}\left[x|\text{no disclosure}\right] > \delta\text{, and therefore } \Gamma_1(\inf(\mathcal{X})+\delta, \hat{y}) > 0$$

(This is true because, as $\hat{y}$ increases, the "no disclosure" region with $x > \delta$ becomes larger and the "no disclosure region with $x < \delta$ becomes smaller.) Consequently, for all $\hat{y} > \inf(\mathcal{Y})+\delta$, $\Gamma_1(\inf(\mathcal{X})+\delta, \hat{y}) > 0$. Analogously, for all $\hat{x} > \inf(\mathcal{X})+\delta$, $\Gamma_2(\hat{x}, \inf(\mathcal{Y})+\delta) > 0$.

Now consider the following claim (stated without proof, as it is analogous to that of Claim 3).

Claim 4. *Let $x^\epsilon = \sup(\mathcal{X}) - \epsilon$ and $y^\epsilon = \sup(\mathcal{Y}) - \epsilon$; then*

$$\lim_{\epsilon\to 0} \mathbb{E}\left[x|\textit{no disclosure region implied by } d_{x_\epsilon, y_\epsilon}\right] < \sup(\mathcal{X}),$$

$$\textit{and } \lim_{\epsilon\to 0} \mathbb{E}\left[y|\textit{no disclosure region implied by } d_{x_\epsilon, y_\epsilon}\right] < \sup(\mathcal{Y}).$$

By Claim 4, we know that there exists some $\delta' > 0$ such that $\Gamma_1(\sup(\mathcal{X})-\delta', \sup(\mathcal{Y})-\delta') < 0$ and $\Gamma_2(\sup(\mathcal{X})-\delta', \sup(\mathcal{Y})-\delta') < 0$. Moreover, we have that for any $\hat{y} < \sup(\mathcal{Y})-\delta'$, it also holds that $\Gamma_1(\sup(\mathcal{X})-\delta', \hat{y}) < 0$. (As $\hat{y}$ decreases, the "no disclosure" region with $x < \delta'$ becomes larger and the "no disclosure region with $x > \delta'$ becomes smaller.) Consequently, for all $\hat{y} < \sup(\mathcal{Y})-\delta'$, $\Gamma_1(\sup(\mathcal{X})-\delta', \hat{y}) < 0$. Analogously, for all $\hat{x} < \sup(\mathcal{X})-\delta'$, $\Gamma_2(\hat{x}, \sup(\mathcal{Y})-\delta) < 0$.

Putting all together, there exists a convex compact subset $\tilde{\mathcal{X}}\times\tilde{\mathcal{Y}}$ of $\mathcal{X}\times\mathcal{Y}$, defined by $\tilde{\mathcal{X}} = [\inf(\mathcal{X})+\delta, \sup(\mathcal{X})-\delta']$ and $\tilde{\mathcal{Y}} = [\inf(\mathcal{Y})+\delta, \sup(\mathcal{Y})-\delta']$ such that for all $\hat{y}\in\tilde{\mathcal{Y}}$,

$$\Gamma_1(\inf(\tilde{\mathcal{X}}), \hat{y}) < 0\text{, and } \Gamma_1(\sup(\tilde{\mathcal{X}}), \hat{y}) > 0.$$

Similarly, for all $\hat{x}\in\tilde{\mathcal{X}}$,

$$\Gamma_2(\hat{x}, \inf(\tilde{\mathcal{Y}})) < 0\text{, and } \Gamma_2(\hat{x}, \sup(\tilde{\mathcal{Y}})) > 0.$$

Therefore, by the Pointcaré-Miranda Theorem, there exist $\hat{x}\in int(\tilde{\mathcal{X}})$ and $\hat{y}\in int(\tilde{\mathcal{Y}})$ such that both $\Gamma_1(\hat{x},\hat{y})$ and $\Gamma_2(\hat{x},\hat{y})$ are simultaneously equal to 0; which are solutions to system (9). ■

Alternative Proof of Statement 2. This is an additional argument showing that full disclosure is not an optimal disclosure rule when p is an affine function. In that case, the optimal disclosure rule must conceal signal realizations with positive probability, and therefore induce $x^{ND}\in \text{int}(\mathcal{X})$ and $y^{ND}\in\text{int}(\mathcal{Y})$. By Theorem 1, $\hat{x} = x^{ND}$ and $\hat{y} = y^{ND}$ must therefore be a solution to system (9).

The demand function is affine, given by $p(x) = a + bx$. For any disclosure rule d, the first term in the seller's payoff expressed in (3) must be $\mathbb{E}(y)\mathbb{E}[P(y,d)] = \mathbb{E}(y)\left[a + b\mathbb{E}(x)\right]$. (This is a straightforward consequence of the martingale property of beliefs, or "Bayesian plausibility.") Therefore the overall probability of sale is independent of the disclosure rule.

Now let d^1 be the full disclosure rule; and take any other disclosure rule $d_{\hat{x},\hat{y}}$, defined by interior thresholds $\bar{x}$ and $\bar{y}$. It must be that, for $y < \bar{y}$, $P(y,d) < P(y,d^1)$ and, for $y > \bar{y}$, $P(y,d) > P(y,d^1)$.

Consequently, the covariance of sales and profitability is larger under d than under d^1; and thus d^1 is not an optimal disclosure rule. ■

Proof of Statement 3. Let $d^{\alpha}_{\hat{x},\hat{y}}$ be the disclosure rule implied by thresholds $\hat{x}$ and $\hat{y}$ when the demand function is given by $p = \alpha p_1 + (1-\alpha)p_2$. Because p_2 is continuously differentiable, regardless of its curvature, the disclosure rule $d^{\alpha}_{\hat{x},\hat{y}}$ induces no disclosure with positive probability, except perhaps if $(\hat{x},\hat{y}) = (\inf(\mathcal{X}), \inf(\mathcal{Y}))$ or $(\hat{x},\hat{y}) = (\sup(\mathcal{X}), \sup(\mathcal{Y}))$.

Moreover, for each $\hat{x} \in int(\mathcal{X})$ and $\hat{y} \in int(\mathcal{Y})$, the values

$$\Gamma_1^{\alpha}(\hat{x},\hat{y}) = \mathbb{E}\left[x|\text{no disclosure region implied by } d^{\alpha}_{x_\epsilon,y_\epsilon}\right] - \hat{x}$$

$$\text{and } \Gamma_2^{\alpha}(\hat{x},\hat{y}) = \mathbb{E}\left[y|\text{no disclosure region implied by } d^{\alpha}_{x_\epsilon,y_\epsilon}\right] - \hat{y}$$

are continuous functions of α. We know by Statement 2 that, for $\alpha = 1$, there is an interior point $(\hat{x},\hat{y})$ such that Γ_1^{α} and Γ_2^{α} are simultaneously equal to 0. By continuity, it must be that there exists $\bar{\alpha} \in (0,1)$ such that if $\alpha > \bar{\alpha}$, then there is an interior point $(\hat{x}',\hat{y}')$ such that Γ_1^{α} and Γ_2^{α} are simultaneously equal to 0 as well.

Proof of Theorem 2

Part 1. Statement of the seller's problem.

Suppose the seller chooses disclosure rule d. Upon observing non-disclosure and profitability y (under mandated transparency), the buyer's mean posterior is

$$x_y^{ND} = \frac{\int_{\mathcal{X}}[1-d(x,y)]x dF_{X|y}(x)}{\int_{\mathcal{X}}[1-d(x,y)] dF_{X|y}(x)}.$$

And so the probability of sale of object y is

$$P(y,d) = \int_{\mathcal{X}} \left[d(x,y)p(x) + (1-d(x,y))p(x_y^{ND})\right] dF_{X|y}(x).$$

Note that the probability of sale of object y is thus independent of the disclosure rule used for objects with profitability $y' \neq y$. And so the seller's problem is separable across profitability levels. Therefore maximizing $\Pi(d)$ over $d: \mathcal{X} \times \mathcal{Y} \to [0,1]$ is equivalent to maximizing, for each $y \in \mathcal{Y}$, $yP(y, d(\cdot,y))$, over $d(\cdot,y): \mathcal{X} \to [0,1]$.

Part 2. Characterization of optimal disclosure rule.

We proceed similarly to the proof of Theorem 1. Fixing some $y \in \mathcal{Y}$ and $x \in \mathcal{X}$, we can take a derivative of the sender's value with respect to $d(x,y)$, to get

$$\frac{\partial \Pi}{\partial d(x,y)} = y\left(p(x)-p(x^{ND})\right) dF_{X|y}(x) dF_Y(y) + \left(\int_{\mathcal{X}} \tilde{y}\left[1-d(\tilde{y},\tilde{x})\right] dF_{X|\tilde{y}}(\tilde{x})\right) p'(x_y^{ND}) \frac{\partial x^{ND}}{\partial d(x,y)}$$

Now from the definition of x_y^{ND}, we get

$$\frac{\partial x_y^{ND}}{\partial d(x,y)} = \frac{\int_{\mathcal{X}}(\tilde{x}-x)(1-d(\tilde{x},y))dF_{X|\tilde{y}}(\tilde{x})}{\left(\int_{\mathcal{X}}(1-d(\tilde{x},y))dF_{X|\tilde{y}}(\tilde{x})\right)^2} dF_{X|y}(x) dF_Y(y)$$

Substituting this into the previous equation, we have

$$\frac{\partial \Pi}{\partial d(x,y)} = y\left[\left(p(x)-p(x_y^{ND})\right)-p'(x_y^{ND})(x-x_y^{ND})\right] dF_{X|y}(x) dF_Y(y). \tag{18}$$

If $\hat{d}$ is a disclosure rule that induces no disclosure with positive probability and does not satisfy (11) and (12) as given in the Theorem, then there is a positive measure of (x,y) such that (18) is strictly positive but $\hat{d}(x,y)<1$ or such that (18) is strictly negative but $\hat{d}(x,y)>0$. This means that $\hat{d}$ can be strictly improved, and cannot be a solution to the seller's problem.

Now we must consider the possibility that the solution is "full disclosure." For each $\hat{y}\in\mathcal{Y}$, and for each $\hat{x}\in\mathcal{X}$, let $d_{\hat{x}}(\cdot,\hat{y})$ be defined by

$$p(x) \geqslant p\left(\hat{x}\right)+p'\left(\hat{x}\right)\left(x-\hat{x}\right) \Rightarrow d_{\hat{x}}(x,\hat{y})=1,$$

$$\text{and } p(x) < p\left(\hat{x}\right)+p'\left(\hat{x}\right)\left(x-\hat{x}\right) \Rightarrow d_{\hat{x}}(x,\hat{y})=0.$$

First suppose for every $\hat{x}\in\mathcal{X}$, $d_{\hat{x}}(\cdot,\hat{y})$ implies that an open interval of realizations in $\mathcal{X}$ are not disclosed. Then the map $\Gamma_{\hat{y}}(\hat{x}) = \mathbb{E}\left(x\middle|\text{ no disclosure region implied by } d_{\hat{x}}(\cdot,\hat{y})\right)$ is a continuous self map in the compact convex set $\mathcal{X}$; and therefore has a fixed point. As in Lemma 1, we can show that the disclosure rule implied by this fixed point is a strict improvement over full disclosure. If instead there exists some $\hat{x}\in\mathcal{X}$ such that $d_{\hat{x}}(\cdot,\hat{y})$ implies almost all realizations are disclosed, then full disclosure is a disclosure rule that satisfies the necessary conditions given by in the Theorem. ■

Proof of Proposition 2

If p is strictly convex, then full disclosure is the only disclosure rule that satisfies the necessary conditions (11) and (12) in Theorem 2. It is therefore the optimal disclosure rule for the seller under mandated transparency; which trivially has more disclosure than the optimal disclosure rule for the seller with hidden motives.

If p is strictly concave, then no disclosure is the only disclosure rule that satisfies the necessary conditions (11) and (12) in Theorem 2. It is therefore the optimal disclosure rule for the seller under mandated transparency; which trivially has less disclosure than the optimal disclosure rule for the seller with hidden motives. ■

Proof of Proposition 3

Proof of Statement 1. If p is locally strictly convex around $\tilde{x}$, there exist $x', x'' \in \mathcal{X}$ with $x' < \tilde{x} < x''$ such that

$$p(x) > p(\tilde{x}) + p'(\tilde{x})(x - \tilde{x}).$$

From Theorem 2, we therefore know that $d_m(x, y) = 1$ for almost all (x, y) with $x \in [x', x'']$. Consequently, $d(x, y) \leqslant d_m(x, y)$ for almost all (x, y) with $x \in (x', x'')$. Further note that, for each $y > y^{ND}$ (the expected profitability of no disclosure implied by d), there exists $\hat{x}(y)$ such that if $x \in (\tilde{x}, \hat{x}(y))$,

$$p(x) < p(\tilde{x}) + \frac{y^{ND}}{y} p'(\tilde{x})(x - \tilde{x}).$$

Therefore, by Theorem 1, almost all such realizations (x, y) are optimally concealed. And therefore there is an open set of such (x, y) with $x \in [x', x'']$ such that $d(x, y) < d_m(x, y)$.

Proof of Statement 2. The proof of the second statement is analogous. ■

Proof of Proposition 4

Suppose (x, y) is such that $(x - \tilde{x})(y - \tilde{y}) > 0$. Then if

$$p(x) > p(\tilde{x}) + \frac{\tilde{y}}{y} p'(\tilde{x})(x - \tilde{x}),$$

it must be that

$$p(x) > p(\tilde{x}) + p'(\tilde{x})(x - \tilde{x})$$

also holds. Therefore, if $(x - \tilde{x})(y - \tilde{y}) > 0$, $d_m(x, y) = 1$ implies $d(x, y) = 1$; and so $d(x, y) \geqslant d_m(x, y)$. If instead (x, y) is such that $(x - \tilde{x})(y - \tilde{y}) < 0$, then

$$p(x) > p(\tilde{x}) + p'(\tilde{x})(x - \tilde{x}) \Rightarrow p(x) > p(\tilde{x}) + \frac{\tilde{y}}{y} p'(\tilde{x})(x - \tilde{x}).$$

Consequently, in that case $d(x, y) = 1$ implies $d_m(x, y) = 1$; and so $d(x, y) \leqslant d_m(x, y)$. ■

Proof of Proposition 5

Assume $y_{min} < 0 < y_{max}$. (The case where all profitabilities are positive or all profitabilities are negative is discussed in the main text.)

Suppose the buyer's mean posterior about the object's value after observing non-disclosure is $\hat{x} \in \mathcal{X}$. Then a seller with profitability $y > 0$, upon observing a signal realization x, will choose to disclose it if and only if $x \geqslant \hat{x}$ (assuming that the seller discloses when indifferent). Conversely, a seller with profitability $y < 0$ will disclose signal realization x if and only if $x \leqslant \hat{x}$. A seller with profitability $y = 0$ is always indifferent between disclosing and not disclosing, and we resolve this indifference with disclosure.

And so Bayesian consistency requires that, in this equilibrium,

$$\hat{x} = \mathbb{E}\left[x|(x-\hat{x})y < 0\right],$$

where we note that the set $\{(x-\bar{x})y < 0\}$ has positive probability, because the joint distribution of profitabilities and signal realizations has full support.

Regarding existence, I remark that there exists some $\hat{x}$ satisfying

$$\hat{x} = \mathbb{E}\left[x|(x-\hat{x})y < 0\right].$$

To see that, note that $\mathbb{E}\left[x|(x-\hat{x})y < 0\right] \in (x_{min}, x_{max})$ for both $\hat{x} = x_{min}$ and $\hat{x} = x_{max}$ (both because the joint distribution of profitabilities and signal realizations has full support). And so, by continuity of $\mathbb{E}\left[x|(x-\hat{x})y < 0\right]$, there must be some solution to $\hat{x} = \mathbb{E}\left[x|(x-\hat{x})y < 0\right]$. Given such $\hat{x}$, it is easy to see that $d^*(x, y) \in \{0, 1\}$ and $d^*(x, y) = 1 \Leftrightarrow (x-\hat{x})y \geqslant 0$ defines an equilibrium disclosure strategy. ■

Proof of Proposition 6

Under mandated transparency, the buyer forms mean posteriors about the object separately for each profitability level y. It follows from a standard unravelling argument (reproduced below) that there can be no concealment in equilibrium.

Suppose for some $y \in \mathcal{Y}$, $\mathbb{X} \subseteq \mathcal{X}$ is the set of signal realizations that the seller does not disclose to the buyer; that is, so that $d(x, y) < 1$ if $x \in \mathbb{X}$. And suppose $\mathbb{X}$ is a set with positive probability according to $F_{X|y}$. Then upon seeing no disclosure, and that the object's profitability is y, the buyer's "no disclosure" belief is

$$x_y^{ND} = \frac{\int_{\mathcal{X}} x(1-d(x,y))f_{X|y}(x)dx}{\int_{\mathcal{X}}(1-d(x,y))f_{X|y}(x)dx}.$$

From this construction, it must be that there is a positive probability that $x \in \mathbb{X}$ and $x > x_y^{ND}$. And therefore, if one such signal x realizes, the seller's best response is to disclose it to the buyer, thereby deviating from the candidate equilibrium strategy $d(x, y) < 1$. We can therefore conclude that, for each y, any $d(\cdot; y)$ such that no disclosure happens with positive probability according to $F_{X|y}$ cannot be an equilibrium disclosure strategy.

And so it follows that the unique equilibrium disclosure strategy is full disclosure. ■

B Numerical Exercise in Sections 3.2.1 and 3.2.2

B.1 Convex Case: Section 3.2.1

We wish to find $\hat{x}$ and $\hat{y}$ such that, given thresholds

$$\bar{x} = \hat{x} \text{ and}$$

$$\bar{y}(x) = \hat{y}\left[\frac{2\hat{x}}{\hat{x}+x}\right],$$

and the disclosure rule as described in Theorem 1 (and depicted in the left-hand panel of Figure 1), these values also satisfy $\hat{x} = \mathbb{E}(x|\text{no disclosure})$ and $\hat{y} = \mathbb{E}(y|\text{no disclosure})$. This problem corresponds to the following system:

$$\hat{x} = A(\hat{x},\hat{y})/B(\hat{x},\hat{y}),$$
$$\hat{y} = C(\hat{x},\hat{y})/D(\hat{x},\hat{y}),$$

where $A(\hat{x},\hat{y})/B(\hat{x},\hat{y})$ is the expected value of x in the no disclosure region for the disclosure rule defined by $\hat{x}$ and $\hat{y}$. Similarly, $C(\hat{x},\hat{y})/D(\hat{x},\hat{y})$ is the expected value of y in the no disclosure region for the disclosure rule defined by $\hat{x}$ and $\hat{y}$. The expressions for each of these objects are

$$\begin{aligned}
A(\hat{x},\hat{y}) &= \int_{\tilde{x}}^{\hat{x}}\left[1-\frac{2\hat{x}\hat{y}}{\hat{x}+x}\right]x\,dx + \int_{\hat{x}}^{1}\frac{2\hat{x}\hat{y}}{\hat{x}+x}x\,dx \\
&= \frac{\hat{x}^2}{2} - \frac{\tilde{x}^2}{2} - 2\hat{x}\hat{y}\left[x - \log(x+\hat{x})\hat{x}\right]_{\tilde{x}}^{\hat{x}} + 2\hat{x}\hat{y}\left[x - \hat{x}\log(x+\hat{x})\hat{x}\right]_{\hat{x}}^{1} \\
&= \frac{\hat{x}^2-\tilde{x}^2}{2} + 2\hat{x}\hat{y}(1+\tilde{x}-2\hat{x}) + 2\hat{x}^2\hat{y}\left[2\log(2\hat{x}) - \log(\tilde{x}+\hat{x}) - \log(1+\hat{x})\right].
\end{aligned}$$

$$B(\hat{x},\hat{y}) = \int_{\tilde{x}}^{\hat{x}}\left[1-\frac{2\hat{x}\hat{y}}{\hat{x}+x}\right]dx + \int_{\hat{x}}^{1}\frac{2\hat{x}\hat{y}}{\hat{x}+x}dx$$

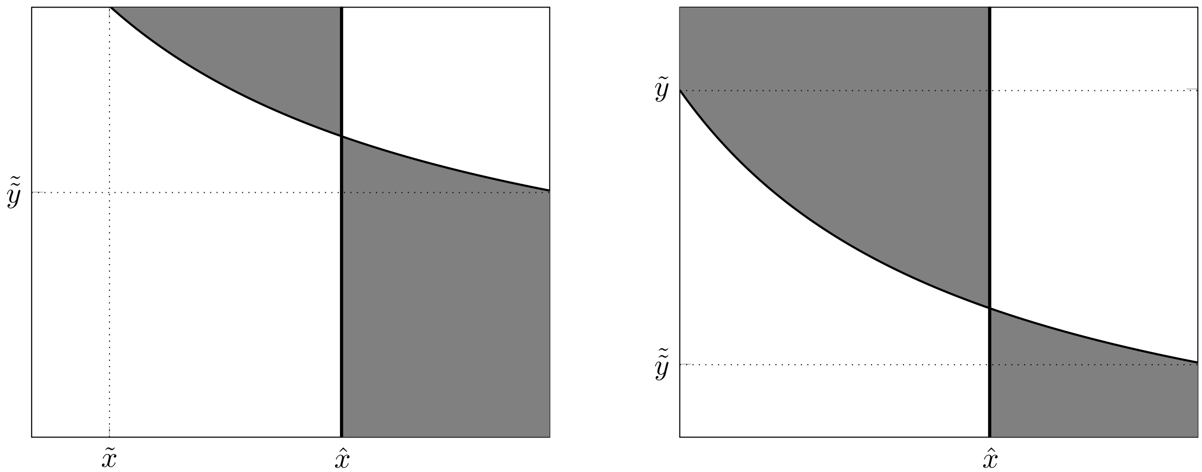

Figure 4: For different values of $\hat{x}$ and $\hat{y}$, each panel illustrates the definition of the objects $\tilde{x}$, $\tilde{y}$, and $\tilde{\tilde{y}}$, used in the calculation of A, B, C, and D in the numerical exercise.

$$= (\hat{x} - \tilde{x}) + 2\hat{x}\hat{y}\left[\log(1+\hat{x}) + \log(\hat{x}+\tilde{x}) - 2\log(2\hat{x})\right].$$

$$C(\hat{x},\hat{y}) = \int_0^{\tilde{\tilde{y}}} (1-\hat{x})y dy + \int_{\tilde{\tilde{y}}}^{\hat{y}} (2\hat{x}\hat{y} - 2\hat{x}y)dy + \int_{\hat{y}}^{\tilde{y}} (2\hat{x}y - 2\hat{x}\hat{y})dy + \int_{\tilde{y}}^{2} \hat{x}y dy$$

$$= \frac{(1-\hat{x})\tilde{\tilde{y}}^2}{2} + \frac{\hat{x}(1-\tilde{y}^2)}{2} + 2\hat{x}\hat{y}(2\hat{y} - \tilde{y} - \tilde{\tilde{y}}) + \hat{x}(\tilde{y}^2 + \tilde{\tilde{y}}^2 - 2\hat{y}^2).$$

$$D(\hat{x},\hat{y}) = \int_0^{\tilde{\tilde{y}}} (1-\hat{x})dy + \int_{\tilde{\tilde{y}}}^{\hat{y}} \left(\frac{2\hat{x}\hat{y}}{y} - 2\hat{x}\right) dy + \int_{\hat{y}}^{\tilde{y}} \left(2\hat{x} - \frac{2\hat{x}\hat{y}}{y}\right) dy + \int_{\tilde{y}}^{2} \hat{x} dy$$

$$= (1-\hat{x})\tilde{\tilde{y}} + \hat{x}(1-\tilde{y}) + 2\hat{x}(\tilde{y} + \tilde{\tilde{y}} - 2\hat{y}) + 2\hat{x}\hat{y}\left[2\log(\hat{y}) - \log(\tilde{y}) - \log(\tilde{\tilde{y}})\right].$$

In the equations above, we use $\tilde{x} = \max\{0, \hat{x}(2\hat{y}-1)\}$, $\tilde{y} = \min\{1, 2\hat{y}\}$, and $\tilde{\tilde{y}} = \frac{2\hat{y}\hat{x}}{1+\hat{x}}$. See Figure 4 for an illustration of the objects $\tilde{x}$, $\tilde{y}$, and $\tilde{\tilde{y}}$. Numerically, I find that the unique $\hat{x}$ and $\hat{y}$ that satisfy the system are $\hat{x} = 0.5824$ and $\hat{y} = 0.5098$. Using these values, we can calculate the expected payoffs to the seller in three benchmarks: full disclosure, no disclosure, and optimal disclosure.

Expected payoff under full disclosure:

$$\frac{1}{2}\int_0^1 x^2 dx = \frac{1}{6}.$$

Expected payoff under no disclosure:

$$\frac{1}{2}\left(\int_0^1 xdx\right)^2 = \frac{1}{8}.$$

Expected payoff under optimal disclosure:

$$\frac{1}{6}+\int_0^{\tilde{\tilde{y}}}\left[\int_{\hat{x}}^1(\hat{x}^2-x^2)dx\right]ydy+\int_{\tilde{\tilde{y}}}^{\hat{y}}\left[\int_{\hat{x}}^{\frac{2\hat{x}\hat{y}}{y}-\hat{x}}(\hat{x}^2-x^2)dx\right]ydy+\int_{\hat{y}}^{\tilde{y}}\left[\int_{2\frac{\hat{x}\hat{y}}{y}-\hat{x}}^{\hat{x}}(\hat{x}^2-x^2)dx\right]ydy$$

$$+\int_{\tilde{y}}^1\left[\int_0^{\hat{x}}(\hat{x}^2-x^2)dx\right]ydy=\frac{1}{6}+\int_0^{\tilde{\tilde{y}}}\left[\hat{x}^2-\frac{2}{3}\hat{x}^3-\frac{1}{3}\right]ydy$$

$$+\int_{\tilde{\tilde{y}}}^{\hat{y}}\left[\frac{2\hat{x}^3\hat{y}}{y}-2\hat{x}^3-\frac{1}{3}\left(\frac{2\hat{x}\hat{y}}{y}-\hat{x}\right)^3+\frac{\hat{x}^3}{3}\right]ydy-\int_{\hat{y}}^{\tilde{y}}\left[\frac{2\hat{x}^3\hat{y}}{y}-2\hat{x}^3-\frac{1}{3}\left(\frac{2\hat{x}\hat{y}}{y}-\hat{x}\right)^3+\frac{\hat{x}^3}{3}\right]ydy$$

$$+\int_{\tilde{y}}^1\frac{2}{3}\hat{x}^3ydy=$$

$$=\frac{1}{6}+\left[\hat{x}^2-\frac{2}{3}\hat{x}^3-\frac{1}{3}\right]\frac{\tilde{\tilde{y}}^2}{2}+\frac{2}{3}\hat{x}^3\left(\frac{1-\tilde{y}^2}{2}\right)+2\hat{x}^3\hat{y}(2\hat{y}-\tilde{\tilde{y}}-\tilde{y})-\frac{5}{3}\hat{x}^3\left[\hat{y}^2-\frac{\tilde{\tilde{y}}^2}{2}-\frac{\tilde{y}^2}{2}\right]$$

$$-\int_{\tilde{\tilde{y}}}^{\hat{y}}\frac{1}{3}\left(2\hat{x}\hat{y}y^2-\hat{x}y^3\right)^3dy+\int_{\hat{y}}^{\tilde{y}}\frac{1}{3}\left(2\hat{x}\hat{y}y^2-\hat{x}y^3\right)^3dy=0.1885$$

B.2 Concave Case: Section 3.2.2

We wish to find $\hat{x}$ and $\hat{y}$ such that, given thresholds

$$\bar{x}=\hat{x}, \text{ and}$$

$$\bar{y}(x)=\hat{y}\left[\frac{2(1-\hat{x})}{2-\hat{y}-x}\right].$$

and the disclosure rule as described in Theorem 1 (and depicted in the right-hand panel of Figure 1), these values also satisfy $\hat{x}=\mathbb{E}(x|\text{no disclosure})$ and $\hat{y}=\mathbb{E}(y|\text{no disclosure})$. This problem corresponds to the following system:

$$\hat{x}=A(\hat{x},\hat{y})/B(\hat{x},\hat{y}),$$
$$\hat{y}=C(\hat{x},\hat{y})/D(\hat{x},\hat{y}),$$

where A, B, C, D are given by the following:

$$A=\int_0^{\hat{x}} xdx-2\hat{y}(1-\hat{x})\int_0^{\hat{x}}\frac{x}{2-\hat{x}-x}dx+2\hat{y}(1-\hat{x})\int_{\hat{x}}^{\tilde{x}}\frac{x}{2-\hat{x}-x}dx+\int_{\tilde{x}}^1 xdx$$

$$=\frac{\hat{x}^2}{2}+\frac{1}{2}-\frac{\tilde{x}^2}{2}+2\hat{y}(1-\hat{x})(2\hat{x}-\tilde{x})+2\hat{y}(1-\hat{x})(2-\hat{x})\left[2\log(2-2\hat{x})-\log(2-\hat{x}-\tilde{x})-\log(2-\hat{x})\right]$$

$$B=\int_0^{\hat{x}}\left[1-\frac{2\hat{y}(1-\hat{x})}{2-\hat{x}-x}\right]dx+\int_{\hat{x}}^{\tilde{x}}\frac{2\hat{y}(1-\hat{x})}{2-\hat{x}-x}dx+\int_{\tilde{x}}^1 1dx$$

$$=\hat{x}+(1-\tilde{x})+2\hat{y}(1-\hat{x})\left[2\log(2-2\hat{x})-\log(2-\hat{x}-\tilde{x})-\log(2-\hat{x})\right]$$

$$C=\int_0^{\hat{y}}(1-\hat{x})ydy+\int_{\hat{y}}^{\tilde{\tilde{y}}}\left[(\hat{x}-1)y+2\hat{y}(1-\hat{x})\right]dy+\int_{\tilde{y}}^{\hat{y}}\left[(2-\hat{x})y-2\hat{y}(1-\hat{x})\right]dy+\int_{\hat{y}}^1\hat{x}ydy$$

$$=(1-\hat{x})\frac{\hat{y}^2}{2}+\hat{x}\left(\frac{1}{2}-\frac{\hat{y}^2}{2}\right)+2\hat{y}(1-\hat{x})\left(\tilde{\tilde{y}}+\tilde{y}-2\hat{y}\right)+(\hat{x}-1)\left(\frac{\tilde{\tilde{y}}^2}{2}-\frac{\hat{y}^2}{2}\right)+(2-\hat{x})\left(\frac{\hat{y}^2}{2}-\frac{\tilde{y}^2}{2}\right).$$

$$D=\int_0^{\hat{y}}(1-\hat{x})dy+\int_{\hat{y}}^{\tilde{\tilde{y}}}\left(\hat{x}-1+\frac{2\hat{y}(1-\hat{x})}{y}\right)dy+\int_{\tilde{y}}^{\hat{y}}\left(2-\hat{x}-\frac{2\hat{y}(1-\hat{x})}{y}\right)dy+\int_{\hat{y}}^1\hat{x}dy$$

$$=\hat{y}(1-\hat{x})+(1-\hat{y})\hat{x}+(\hat{x}-1)(\tilde{\tilde{y}}-\hat{y})+(2-\hat{x})(\hat{y}-\tilde{y})+2\hat{y}(1-\hat{x})\left(\log\tilde{\tilde{y}}+\log\tilde{y}-2\log\hat{y}\right).$$

In the equations above, we use $\tilde{x}=\min\{1,\hat{x}+2-2\hat{y}(1-\hat{x})\}$, $\tilde{y}=\frac{2\hat{y}(1-\hat{x})}{2-\hat{x}}$, and $\tilde{\tilde{y}}=\min\left\{1,\frac{2\hat{y}(1-\hat{x})}{1-\hat{x}}\right\}$. Numerically, I find that the unique $\hat{x}$ and $\hat{y}$ that satisfy the system are $\hat{x}=0.5397$ and $\hat{y}=0.5402$. Using these values, we can calculate the expected payoffs to the seller in three benchmarks: full disclosure, no disclosure, and optimal disclosure.

Expected payoff under full disclosure:

$$\frac{1}{2}\int_0^1(2x-x^2)dx=\frac{1}{3}.$$

Expected payoff under no disclosure:

$$\frac{1}{2}\left[2\left(\int_0^1 xdx\right)-\left(\int_0^1 xdx\right)^2\right]=\frac{3}{8}.$$

Expected payoff under optimal disclosure:

$$\frac{1}{2}\left(2\hat{x}-\hat{x}^2\right)+\int_0^{\tilde{y}}\left[\int_0^{\hat{x}}\left(2(x-\hat{x})-(x^2-\hat{x}^2)\right)dx\right]ydy$$

$$+\int_{\tilde{y}}^{\hat{y}}\left[\int_{2-\hat{x}-\frac{2\hat{y}(1-\hat{x})}{y}}^{\hat{x}}\left(2(x-\hat{x})-(x^2-\hat{x}^2)\right)dx\right]ydy$$

$$+\int_{\hat{y}}^{\tilde{\tilde{y}}}\left[\int_{\hat{x}}^{2-\hat{x}-\frac{2\hat{y}(1-\hat{x})}{y}}\left(2(x-\hat{x})-(x^2-\hat{x}^2)\right)dx\right]ydy+\int_{\tilde{\tilde{y}}}^{1}\left[\int_{\hat{x}}^{1}\left(2(x-\hat{x})-(x^2-\hat{x}^2)\right)dx\right]ydy$$

$$=\frac{1}{2}\left(2\hat{x}-\hat{x}^2\right)+\int_0^{\tilde{y}}\left(\frac{2}{3}\hat{x}^3-\hat{x}^2\right)ydy+\int_{\tilde{\tilde{y}}}^{1}\left[-\frac{2}{3}\hat{x}^3+2\hat{x}^2-2\hat{x}+\frac{2}{3}\right]ydy$$

$$+\int_{\tilde{y}}^{\hat{y}}\left[(\hat{x}^2-2\hat{x})\left(2\hat{x}-2+\frac{2\hat{y}(1-\hat{x})}{y}\right)+\hat{x}^2-\left(2-\hat{x}-\frac{2\hat{y}(1-\hat{x})}{y}\right)^2-\frac{\hat{x}^3}{3}\right.$$

$$\left.+\frac{1}{3}\left(2-\hat{x}-\frac{2\hat{y}(1-\hat{x})}{y}\right)^3\right]ydy-\int_{\hat{y}}^{\tilde{\tilde{y}}}\left[(\hat{x}^2-2\hat{x})\left(2\hat{x}-2+\frac{2\hat{y}(1-\hat{x})}{y}\right)+\hat{x}^2\right.$$

$$\left.-\left(2-\hat{x}-\frac{2\hat{y}(1-\hat{x})}{y}\right)^2-\frac{\hat{x}^3}{3}+\frac{1}{3}\left(2-\hat{x}-\frac{2\hat{y}(1-\hat{x})}{y}\right)^3\right]ydy=0.3907.$$