Sequential Screening via Trial Versions*

Click here for the latest version.

Yihu Hou

December 9, 2022

Abstract

Quantity of information is an essential screening tool yet in reality we see a stark dichotomy: some firms offer various trial sizes while others offer a homogeneous one. To find out why this dichotomy occurs, I model the scenario when a firm can use trials to persuade consumers, who do not fully know their preference, into experimenting and subsequently purchasing the full upgrade. The model differs from traditional ones in that learning is only possible through the consumption of the sample, which limits the information provision and rent extraction of the principal. I find that the firm's information provision depends on the way in which consumers obtain information from consuming the sample. The optimal screening mechanism with bad news belief updating involves full bunching (i.e. a single trial version), while good news belief updating, on the other hand, leads to full discrimination (i.e. a rich trial menu). In addition, the paper complements the traditional method in dynamic screening by providing a full characterization of the solution when local incentive condition does not extend globally and when the "double-deviation" i.e., off-path misreporting occurs. I show these requirements lead to more bunching.

*I am very grateful for the guidance from Andrew Rhodes, Bruno Jullien, and Jacques Crémer. I also thank Alexander Guembel, Fabrice Collard, Alexandre de Cornière, and François Salanié for their comments. I gratefully acknowledges financial support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 759733), and from Takuro Yamashita's ERC grant 714693.

1 Introduction

Trial versions are everywhere. With lots of software, we can start with a free or cheap version and later pay additional fees to unlock more comprehensive and specialized features. In cosmetic stores, we often see a variety of perfume samples, which can help sellers to market their products. Trial versions also come in the form of time limits. For example, we can get a discount for the first month of membership on a music streaming platform, but after that, we have to pay the full price. A trial version can be a version with limited functionality, duration, quantity, or quality.

Despite the prevalence of trial versions, the economics of trial versions has not been well studied. In general, a larger trial version provides consumers with more precise information. Thus the firm can screen consumers in terms of information by offering a rich menu of different sizes of trials. A typical puzzle is why some companies offer only a single trial version, while some other companies offer rich trial options. This distinction cannot be well predicted by existing theories of information design.

For example, we can start a mobile plan with a new customer offer, which provides a discount with a limited duration. We can learn how the mobile plan works in our regions of activities in the trial period. The longer we try, the more we learn. However, the trial period is usually the same for everyone, which means that the mobile network operator is not trying to discriminate against different consumers in terms of information. An opposite example is video games. Usually, when a game company releases a new game, it gives players a first top-up offer, which can be regarded as a form of a trial version. The first top-up offer is usually tiered, such as getting 200 for 100 paid, 320 for 200, 700 for 500, and so on. Players who are optimistic about this new game usually go straight for the bigger top-up options, while those who are not so confident will choose the smaller options.

What explains the difference between the trial versions of the video game and the mobile plan? In this paper, I found that different information structures led sellers to adopt different trial options. We obtain new information and update our beliefs about the goods through trial versions. New information can be released in the form of good news or bad news. If the belief is updated by good news, when the agent receives good news he confirms the product is of high valuation to him, and when he does not receive any good news, he is unsure whether he is of high or low valuation. If the belief is updated by bad news, when the agent receives bad news he confirms the product is of low valuation to him, and when he does not receive any bad news, he is unsure whether he is of high or low valuation. For example, with mobile plans, information is released more often in the form of bad news. As the users try a mobile plan, at first it seems to work well with a good internet connection. But a day may come when the internet connection breaks down or the call quality becomes poor. That is where the bad news occurs. Video games, on the other hand, are a very different kind of belief update, where the information is released more in the form of good news. Users start playing a video game and the first level of the game makes them feel bored. But if they keep playing, they may find it very exciting from a certain level onwards, and then they get obsessed. That is where the good news occurs. In this paper, I found that, in good news belief updates, companies have the incentive to provide rich trial menus, while in bad news belief updates, companies only want to provide a single trial version[1].

In addition to the applied result, this paper has theoretical contributions. In dynamic screening problems, the first-order local approach often fails. Existing literature deals with this by introducing conditions under which the first-order local approach works[2]. But this has led to a serious selection bias - restricting our attention only to the settings where the standard approach works has biased our understanding of information screening a lot, because the standard approach usually works in settings where discrimination is easy to implement. Thus, some important insights are actually missing. There is no way to actually deal with non-local imitation especially when it comes to double deviation (off-path misreporting), and we do not know yet how this is going to reshape the optimal mechanism. In this paper, I complement the traditional method by providing a full characterization of the solution when the local incentive condition does not extend globally and when the "double-deviation" i.e., off-path misreporting occurs. I show these requirements lead to more bunching. Best to the author's knowledge, this is the first paper to provide an explicit solution that is immune to an otherwise binding double-deviation constraint. Thus, a brand new insight arises in this paper, that is, how a firm screens by information provision depends crucially on the information structure - how consumers learn about their valuations.

One trade-off for the firm is that, on the one hand, they want consumers to have some information so they are willing to pay higher prices, and on the other hand, they do not want consumers to learn bad information. This contradiction is not so intense in the good news belief updating, because the new information comes in the form of good news. But

[1]Some products may be a mix of good news and bad news belief update. However, the two distinct models are enough to capture the main sight and help us understand how different information structures affect a firm's optimal selling strategy.

[2]The usual setting is that the products are indivisible or with a specific information structure (e.g. prior and new information are independently additive). Courty and Li (2000) points out that with divisible quantities, generally we fail to characterize the optimal mechanism because the first-order condition is not sufficient.

with bad news belief updating, this contradicts a lot. This will result in the firm wanting to provide less information to a consumer who is already an optimistic ex-ante. Thus, bunching happens in the bad news belief updating because of the dual role of the trial version: it is both a consumption and information. In the classical mechanism design, the more optimistic the consumer is, the more quantity he or she consumes. However, if consumers learn by trials, with bad news belief updating, an optimistic consumer tends to need less information, which means he or she needs less quantity in the trial version. This leads to an inverted quantity allocation, which further gives room for non-local imitation and double deviation. To deter those irregular deviations, the firms have to bunch more.

In the following of the paper, section 3 presents the setting of the model. Section 4 analyzes the good news model and section 5 analyzes the bad news model. Section 6 explains the intuition of the difference between the good and bad news model. Section 7 gives further implications.

2 Literature

2.1 Sequential Screening

The paper contributes directly to sequential screening, a sub-field of dynamic mechanism design. Courty and Li (2000) is the first to study the situation where the consumers are initially uninformed about their valuations, but only the distribution of it, which is the *ex-ante* type of each consumer. The optimal mechanism takes the form of refund contracts. This paper differs from Courty and Li (2000) in a few aspects: i) Courty and Li (2000) assumes a unit demand while here quantity is an important screening tool, which also determines the amount of learning; ii) learning is automatic in theirs but endogenously chosen by the principal here. Therefore, I provide an additional scope, which illustrates the tension between rent extraction in the learning stage and ex-post. Eső and Szentes (2007a) and Eső and Szentes (2007b) show a dynamic screening model can be transformed into a setting where the buyer acquires an *independent* piece of information, i.e., an orthogonal transformation. Li and Shi (2017) studies the case where the release of information is controlled by the sender and find that i) sender may want to release garbled information, which echos my solution where agents are offered limited sample sizes, which determines the limited amount of information learned by the agents; ii) orthogonal transformation is invalid when disclosure or learning is part of the mechanism, which justifies my use of a specific information structure. Bergemann and Välimäki (2019) provides an up-to-date survey of dynamic mechanism design.

A few papers, including but not limited to Baron and Besanko (1984), Battaglini (2005), Pavan et al. (2014) and Boleslavsky and Said (2013), study the dynamic mechanism design problem using the first-order approach (i.e., only local incentive compatibility is considered), which is usually only true under a regularity condition, which, as pointed out by Battaglini and Lamba (2019), is not easy to satisfy. Indeed, I show the optimal contracts not prone to a global deviation involve more bunching than the benchmark case in which only local incentive conditions are considered.

Another subtlety in the paper is "double-deviation" that agents misreport their types in both stages. The dynamic revelation principle, in general, cannot guarantee off-path truth-telling. The issue was well-aware, for example, in Börgers and Krahmer (2015) and Bergemann and Välimäki (2019). The requirement for the mechanism to be immune to double-deviation is irrelevant in Courty and Li (2000) as their *ex-post* type is their payoff type, so truth-telling is always incentive compatible even off the equilibrium path. Here, it is not the case. Offering samples is costly, the principal may have to subsidize the buyers for them to discover their true tastes by offering a cheaper sample. But then, other agents with lower ex-ante types may want to exploit the cheaper sample and run away in the second period. To deter this type of double deviation, the offer has to involve too much bunching.

Some recent studies, including Wei and Green (2019) and Guo et al. (2022), have found that sellers give different amounts of information to different consumers in order to discriminate, without getting bunching results. This is because i) Wei and Green (2019) set up in such a way that information can be released directly and agents learn before they consume, so giving more information to pessimistic consumers does not cause incentive compatibility problems; ii) Wei and Green (2019) and Guo et al. (2022) both analyze the case where the prior information and the ex-post information are independent and additive. Their results are similar to my good news model, i.e., when prior information and ex-post information are decomposable in some sense (no matter whether they are independently additive or ex-post information directly equals posterior belief), we do not have the non-local imitation and double deviation problems. Without these additional deviation problems, it will be easier for sellers to discriminate in terms of information, so there will be no bunching.

2.2 Bayesian Persuasion

This paper is also related to the Bayesian Persuasion literature, stemming from the seminal paper by Kamenica and Gentzkow (2011). Most papers[3] in this area study arbitrary signals, only subject to a martingale constraint that the expectation of the posterior has to equal the prior. A few notable exceptions include: Gentzkow and Kamenica (2014) assumes the cost of signals is proportional to the uncertainty reduction; In Nguyen and Tan (2021), the sender communicates a message whose cost depends on both the message content and the signal realization; Matyskova (2018) instead looks at the case where the receiver has other sources of costly information. The cost of information is independent of the receiver in all of these models. Here, the sample serves as a signal, which the seller calibrates to manipulate the buyer's belief and persuades them to buy. However, it is also costly even if samples have no production costs: the seller has to restrict the rent he extracts from the sample in order to balance the persuasion. An optimal schedule trades off the amount of information learned by the buyers and the rent from selling the sample.

2.3 Applied Study

There are also some papers on experience goods, including Bergemann and Välimäki (2006). They study indivisible products in a dynamic setting and are not about screening as there is only a single price per period.

There are also several management papers that discuss freemium and versioning, including Bhargava and Choudhary (2008), Chellappa and Mehra (2018), Chellappa and Shivendu (2005), etc. They study some specific problems, such as how the trial version affects piracy, how to use freemium to correct consumers' biased prior beliefs, and so on.

3 Model

A monopoly sells a divisible product, and each consumer can consume up to quantity Q (the full product). Consumers have two possible valuations of this product, high or low. The consumer utility is linear in quantity. The consumer with a high valuation receives Q utility from the full product, while the consumer with a low valuation has a utility of 0. Consumers do not know whether they are of high or low valuation ex-ante, but they know the probability x that they are of high valuation. The prior x is the consumer's private

[3] See Kamenica (2019) for a comprehensive survey.

information and the firm believes that x follows a distribution F.[4]

This setting focuses more on horizontal differentiation (the valuation is idiosyncratic in that different consumers have different valuations of the same product) than vertical differentiation although it allows for the vertical one technically[5]. Vertical differentiation, like the objective quality of the product, can be learned from many other ways such as a review system, third-party recommendations/certification, and word-of-mouth. But horizontal differentiation, the idiosyncratic matching between the consumer and the product usually can only be learned through real consumption.

The firm sells its products in two stages. It offers consumers a menu, where each option includes a trial size q_1, the corresponding trial price p_1, and the corresponding unit price p_2 for continued upgrades after the trial. Consumers can choose from those options.

Consumers may get a new signal[6] while using the trial version, and the signal is not observable to the firm. The larger the trial version, the higher the probability of obtaining a new signal. I will consider two different specifications of the information generation process, the good news model and the bad news model. In the good news model, the signal arrives only if the consumer is of high valuation so the consumer who receives the signal can confirm that he has a high valuation. The low-valuation customer never receives the signal in the good news model, no matter how large his trial size is. In the bad news model, the signal arrives only if the consumer is of low valuation so the consumer who receives the signal can confirm that he has a low valuation. The high-valuation customer never receives the signal in the bad news model, no matter how large his trial size is.

In the good news model, the probability that consumer x receives the signal (i.e., the good news) if he purchases and consumers trial version of size q_1, is the ex-ante probability that he is of high valuation x multiplied by the probability that the good news lands in the trial version $G(\frac{q_1}{Q})$, i.e., $xG(\frac{q_1}{Q})$, where G is a cumulative distribution function from 0 to 1. Similarly in the bad news model, the probability that consumer x receives bad news if he purchases and consumers the trial version of size q_1 is the probability that he is of low valuation $1 - x$ multiplied by the probability that the bad news lands in the trial version $G(\frac{q_1}{Q})$, i.e., $(1 - x)G(\frac{q_1}{Q})$.

In the good news model, a consumer with a high valuation must receive good news in the process of consuming the full product. In the bad news model, a consumer with a low valuation must receive bad news in the process of consuming the full product, i.e.

[4]F can also be common knowledge, but it does not make any differences because F is not considered in the consumer's decision.

[5]The vertical differentiation can be obtained by shifting the distribution function $F(x)$ as a whole.

[6]The consumption utility of the trial depends on the prior and the signal, so there is no extra information besides the signal.

$G(\frac{Q}{Q}) = G(1) = 1$. This is consistent with the fact that G is a cumulative distribution function from 0 to 1. Even if consumers do not receive any signal in the trial version, they will update their beliefs, because not receiving good news can lead to bad belief updates, and not receiving bad news can lead to good belief updates. Whether in the good news model or in the bad news model, the scenarios are binary. I define the good scenario as the scenario where the posterior belief is higher than the ex-ante prior, while the bad scenario is the scenario where the posterior belief is lower than the ex-ante prior. In the good news model, receiving good news is the good scenario, while not receiving good news is the bad scenario. In the bad news model, receiving bad news is the bad scenario, while not receiving bad news is the good scenario.

We can calculate the consumer's expected utility per unit of the goods conditional on the good scenario. In the good news model, it is 1, because once a consumer receives the good news, he confirms that he is of high valuation. In the bad news model, it is

$$Prob(\text{ high } | \text{ no bad news }) = \frac{Prob(\text{ no bad news } | \text{ high })Prob(\text{ high })}{Prob(\text{ no bad news })} = \frac{1 \cdot x}{1 - (1 - x)G(\frac{q_1}{Q})}.$$

Accordingly, we can calculate the consumer's expected utility per unit of the goods conditional on the bad scenario. In the bad news model, it is 0, because once a consumer receives the bad news, he confirms that he is of low valuation. In the good news model, it is

$$Prob(\text{ high } | \text{ no good news }) = \frac{Prob(\text{ no good news } | \text{ high })Prob(\text{ high })}{Prob(\text{ no good news })} = \frac{[1 - G(\frac{q_1}{Q})] \cdot x}{1 - xG(\frac{q_1}{Q})}.$$

The firm's mechanism can be written as

$$\{q_1(x), p_1(x), p_2(x)\},$$

where x is the index of different options in the menu, which at the same time, indicates that it is the option chosen by consumer x.

I now impose regularity conditions for $F(\cdot)$ and $G(\cdot)$.

Assumption 1. $\frac{xf(x)}{1-F(x)}$ *increases in* x.

This is an assumption commonly used in mechanism design. It is weaker than the "increasing hazard rate" and satisfied by most commonly used distributions. This condition says that conditional on the type is at least x, the conditional expected marginal type is increasing in x.

Define $q_2 = Q - q_1$ and

$$M(\frac{q_2}{Q}) = 1 - G(\frac{q_1}{Q}) = 1 - G(1 - \frac{q_2}{Q}).$$

The M function is used to represent the relationship between the remaining quantity and the remaining information apart from the sample.

Assumption 2. *The q_2-elasticity of $M(\frac{q_2}{Q})$ weakly increases in q_2.*

This assumption implies that the preceding unit is more informative than the following unit in the commodity. For example, G is a uniform distribution if the information is evenly distributed among the goods. Then we get that M is also a uniform distribution, and the elasticity is constant at 1. In general, a concave G easily satisfies this condition. However, this condition does not require that G must be concave. As an example, if $G = 1-(1-\frac{q_1}{Q})^{\frac{1}{n}}$, G is convex for $n > 1$ but the elasticity of M is constant at $\frac{1}{n}$, which means this assumption is still satisfied.

Next, I will discuss the revelation principle, that is, the rewriting of any incentive-compatible mechanism into a more tractable form. The "revelation principle" argument implies that it is without loss of generality to assume that a signal comes together with an "action recommendation"[7], and the agents are going to obey the recommendations. In the following proposition, I show that any incentive-compatible mechanism can be rewritten as an incentive-compatible mechanism where, for consumer x, the recommended action in the good scenario is to upgrade up to $q(x)$ and in the bad scenario is not to upgrade.

Proposition 1. *Any incentive-compatible mechanism $\{\tilde{q}_1(x), \tilde{p}_1(x), \tilde{p}_2(x)\}$ can be rewritten as an incentive-compatible mechanism $\{q_1(x), p_1(x), p_2(x), q(x)\}$ where $q(x) \in \{q_1(x), Q\}$ and the consumer x upgrades and consumes up to $q(x)$ in the good scenario and stays at $q_1(x)$ in the bad scenario.*

Proof. Because the utility is linear with respect to quantity, in the second stage, consumers will either buy the entire goods or they will not buy any upgrades.

If the original mechanism $\{\tilde{q}_1(x), \tilde{p}_1(x), \tilde{p}_2(x)\}$ has consumers upgrading in and only in the good scenario, then we can naturally rewrite it as $\{\tilde{q}_1(x), \tilde{p}_1(x), \tilde{p}_2(x), q(x)\}$ by letting $q(x) = Q$ for any x such that $q_1(x) > 0$.

If the original mechanism $\{\tilde{q}_1(x), \tilde{p}_1(x), \tilde{p}_2(x)\}$ has some consumer x that upgrades regardless of the signal, then we can rewrite it as $\{q_1(x), p_1(x), p_2(x), q(x)\}$ by letting

$$q_1(x) = Q, \quad p_1(x) = \tilde{p}_1(x) + (Q - \tilde{q}_1(x))\tilde{p}_2(x), \quad q(x) = Q,$$

[7] Bergemann and Morris (2019).

and $p_2(x)$ can be chosen arbitrarily, weakly less than the consumer's expected utility per unit of the goods conditional on the good scenario, and greater than that conditional on the bad scenario. In this new mechanism, the consumer x will upgrade in and only in the good scenario, but the upgrade package is empty. The utility of consumer x remains the same since it makes no difference whether one consumes a trial version of size Q directly or a trial version of $\tilde{q}(x)$ followed by an upgraded version of $Q - \tilde{q}(x)$. The incentive for others to imitate the consumer x is also weakened because they lose the opportunity to stop at the middle $\tilde{q}(x)$. So if the original mechanism is incentive compatible, then this new mechanism is also incentive compatible.

If the original mechanism $\{\tilde{q}_1(x), \tilde{p}_1(x), \tilde{p}_2(x)\}$ has consumers x that do not upgrade no matter what signal they receive, then we can rewrite it as $\{\tilde{q}_1(x), \tilde{p}_1(x), p_2(x), q(x)\}$ by letting $q(x) = \tilde{q}_1(x)$ and $p_2(x)$ can be chosen arbitrarily, weakly less than the consumer's expected utility per unit of the goods conditional on the good scenario, and greater than that conditional on the bad scenario. Same as what was said earlier, in this new mechanism, the consumer x will upgrade in and only in the good scenario, but the upgrade package is empty. The utility of consumer x remains the same. The incentive for others to imitate the consumer x is also weakened because they lose the opportunity to consume more than $\tilde{q}_1(x)$. So if the original mechanism is incentive compatible, then this new mechanism is also incentive compatible. □

Before the transformation, because consumers do not necessarily upgrade if and only if they are in a good scenario, we cannot simply write the second stage IC with respect to the upgrade price. However, this proposition does the transformation that any incentive-compatible mechanism can be rewritten as an incentive-compatible mechanism that the consumer will upgrade if and only if the scenario is good. This allows us to write the simple and direct second-stage IC condition, which is, in the good news model,

$$\frac{[1 - G(\frac{q_1(x)}{Q})]x}{1 - xG(\frac{q_1(x)}{Q})} < p_2(x) \leq 1,$$

and in the bad news model,

$$0 < p_2(x) \leq \frac{x}{1 - (1 - x)G(\frac{q_1(x)}{Q})}.$$

In the dynamic mechanism design problem, we need to consider the possibility of double deviation, i.e., off-path misreporting. A consumer can imitate someone else in the first stage and misrepresent the signal he received in the second stage. In the good news model, we may find this double deviation in the downward imitation. If an optimistic consumer buys a pessimistic consumer's plan, it is possible that the optimistic consumer will upgrade even when he does not receive the good news. In the bad news model, on the contrary, we may find this double deviation in the upward imitation. When a pessimistic consumer buys the plan of an optimistic consumer, he may not upgrade even if he does not receive the bad news.

To deal with this issue, I first assume that the consumers cannot double deviate and then solve the relaxed problem. So, in the following sections, please distinguish between the original problem and the relaxed problem. We will see later that in the good news model the optimal solution of the relaxed problem does not provide the consumer with an incentive to double deviate, so the original problem is solved in the same way as the relaxed problem. However, in the bad news model, double deviation sometimes occurs.

Define $R(x)$ as the rent of consumer x. In the good news model,

$$R(x) = [q_1(x)x - p_1(x)] + xG(\frac{q_1(x)}{Q})[q(x) - q_1(x)][1 - p_2(x)],$$

and in the bad news model,

$$\begin{aligned} R(x) &= [q_1(x)x - p_1(x)] + \left[1 - (1-x)G(\frac{q_1(x)}{Q})\right][q(x) - q_1(x)]\left[\frac{x}{1 - (1-x)G(\frac{q_1(x)}{Q})} - p_2(x)\right] \\ &= q(x)x - p_1(x) - p_2(x)[q(x) - q_1(x)]\left[1 - (1-x)G(\frac{q(x)}{Q})\right]. \end{aligned}$$

Lemma 1. *The necessary conditions of the first-stage IC condition are*

$$R'(x) = q_1(x) + G(\frac{q_1(x)}{Q})[q(x) - q_1(x)][1 - p_2(x)]$$

in the good news model, and

$$R'(x) = q(x) - p_2(x)G(\frac{q_1(x)}{Q})[q(x) - q_1(x)]$$

in the bad news model, and $R'(x)$ must be non-decreasing in both models. These conditions are sufficient in the relaxed problem without considering double deviation.

See Appendix A for detailed proof. The expression for $R'(x)$ is obtained by the standard first-order approach (envelope theorem), that is, each consumer does not imitate the consumer to his adjacent left. Define $r(x, x')$ as the rent where the consumer x imitates

$x' < x$, and we can get $R'(x)$ by the envelope theorem[8],

$$R'(x) = \frac{\partial r(x, x')}{\partial x}\bigg|_{x' \to x}.$$

Next, we will look at how $p_2(x)$ affects revenue. Higher upgrade prices can help reduce $R'(x)$ in both the good news model and the bad news model. Consider the following case, if a consumer has a 50-50 chance of upgrading, and the firm charges 10 cents less for the upgrade and 5 cents more for the trial version, it does not appear to have any impact on the firm's profits. This is because we have only focused on one individual consumer. However, it will indeed influence the firm's profits through the channel of affecting rent difference between different consumers. Ex-ante more optimistic consumers have the advantage that they are more likely to be of high valuation and more likely to be in a good scenario. If we charge a higher price for the upgrade, we erode this advantage for them: even if they are of high valuation, they will have to pay a high price for additional consumption. A high upgrade price can make consumers in a sense more homogeneous so that the firm can better drain the consumer surplus. This also tells us how the company should distribute the consumer's rent between the two stages. The upgrade price, as an instrument that regulates the two-stage rent allocation, affects the firm's revenue. In addition to this, in section 5.4 we will see that it also affects incentive compatibility in the bad news model.

4 Good News Model

With the good news belief updating, the following is the relaxed problem without considering double deviation.

$$\max \int_0^1 \left[q_1(x)x + xG(\frac{q_1(x)}{Q})[q(x) - q_1(x)] - R(x) \right] f(x)dx$$

[8]This expression for $R'(x)$ is different from what we usually see. According to the revenue equivalence theorem, the $R'(x)$ expression we usually see contains only the quantity allocation terms, but not the payment terms. But here we have $p_2(x)$. In fact, this is not contradictory if we rewrite $p_2(x)$ as a function of quantities $\min_\theta\{q(x, \theta) = Q\}$, where θ is off-path or on-path posterior beliefs. So we should recognize that off-path beliefs are also important in this problem.

$$R'(x) = q_1(x) + G(\frac{q_1(x)}{Q})[q(x) - q_1(x)][1 - p_2(x)] \qquad \text{(First-order First-stage IC)}$$

$$R'(x) \text{ is non-decreasing in } x \qquad \text{(Second-order First-stage IC)}$$

$$\frac{[1 - G(\frac{q_1(x)}{Q})]x}{1 - xG(\frac{q_1(x)}{Q})} < p_2(x) \leq 1, \forall x \qquad \text{(Second-stage IC)}$$

$$R(x) \geq 0, \forall x \qquad \text{(IR)}$$

$$q(x) \in \{q_1(x), Q\}, \forall x \qquad \text{(Proposition 1)}$$

The higher upgrade price $p_2(x)$, the lower $R'(x)$, and the better profits for the firm. Let the condition $p_2(x) \leq 1$ bind, and we have $R'(x) = q_1(x)$. Because $R'(x) \geq 0$, we have $R(0) = 0$ and $R(x) = \int_0^x R'(t)dt$. By integration by parts, the objective function can be rewritten as

$$\max \int_0^1 \left[q_1(x)x + xG(\frac{q_1(x)}{Q})[q(x) - q_1(x)] \right] f(x) - q_1(x)(1 - F(x))dx.$$

Because the objective function increases in $q(x)$, easy to get $q(x) = Q$ for all x.

Next, I will solve for the optimal trial size $q_1(x)$. Define

$$V(q_1, x) = \left[q_1 x + xG(\frac{q_1}{Q})(Q - q_1) \right] f(x) - q_1(1 - F(x)),$$

and thus the optimal trial size for consumer x is the q_1 that maximizes $V(q_1, x)$.

$$\frac{\partial V(q_1, x)}{\partial q_1} \geq 0 \quad \Leftrightarrow \quad 1 + g(\frac{q_1}{Q}) \left(1 - \frac{q_1}{Q} \right) - G(\frac{q_1}{Q}) \geq \frac{1 - F(x)}{xf(x)} \tag{1}$$

In the inequality (1), the left side represents the motivation to increase the trial size and the right side represents the motivation to decrease the trial size. Under the assumption 1, when x increases, the right side decreases, so the motivation to decrease the trial size will decrease and will lead to a larger trial version. See Appendix B for more rigorous proof.

Once $q_1(x)$ is increasing with x, the condition that $R'(x) = q_1(x)$ is non-decreasing is also satisfied. Further, we also rule out the possibility of double deviation since $p_2(x) = 1$, which means that consumers will upgrade if and only if they receive the good news. So this solution is the optimal solution to the original problem.

Proposition 2. *In the good news model, the unit price of the upgrade is 1, and the size of the trial* *version increases in ex-ante type x.*

From this proposition we can see that the firm has set the price of the upgrade very high, leaving the consumer with no rent at all in the second stage. All the rents are offered in the first-stage trial version.

The firm offers a rich menu of different trial sizes. More optimistic consumers will consume larger samples. On the one hand, the unit price of samples is cheaper than the unit price of upgrades, so optimistic consumers are willing to consume a larger sample; on the other hand, the firm is willing to offer a larger trial to optimistic consumers because a larger trial is more likely to bring them the good news and raise their willingness to pay.

5 Bad News Model

With the bad news belief updating, the following is the relaxed problem without considering double deviation.

$$\max \int_0^1 [q(x)x - R(x)]f(x)dx$$

such that

$$R'(x) = q(x) - p_2(x)G(\frac{q_1(x)}{Q})[q(x) - q_1(x)] \quad \text{(First-order First-stage IC)}$$

$$R'(x) \text{ is non-decreasing in } x \quad \text{(Second-order First-stage IC)}$$

$$0 < p_2(x) \leq x\left[1 - (1-x)G(\frac{q_1(x)}{Q})\right]^{-1}, \forall x \quad \text{(Second-stage IC)}$$

$$R(x) \geq 0, \forall x \quad \text{(IR)}$$

$$q(x) \in \{q_1(x), Q\}, \forall x \quad \text{(Proposition 1)}$$

Now I proceed to show that $R'(x) \geq 0$ so that the IR conditions can be reduced to $R(0) \geq 0$. Note, from the second-stage IC we know

$$p_2(x)G(\frac{q_1(x)}{Q}) \leq \frac{xG(\frac{q_1(x)}{Q})}{1-(1-x)G(\frac{q_1(x)}{Q})} \leq 1,$$

and substitute it into the envelope condition we get

$$R'(x) \geq q(x) - 1 \times [q(x) - q_1(x)] = q_1(x) \geq 0.$$

Therefore, I can set the rent of the lowest type to 0, i.e., $R(0) = 0$ and express the rent in the form $R(x) = \int_0^x R'(t)dt$. By integration by parts,

$$\max \int_0^1 [q(x)xf(x) - R'(x)(1 - F(x))]dx$$

$$s.t. \quad R'(x) = q(x) - p_2(x)G(\frac{q_1(x)}{Q})[q(x) - q_1(x)] \tag{2}$$

$$R'(x) \text{ is non-decreasing in } x$$

$$p_2(x) \leq \frac{x}{1 - (1 - x)G(\frac{q_1(x)}{Q})} \tag{3}$$

$$q(x) \in \{q_1(x), Q\}$$

Condition (2) is what we usually call the first-order IC condition. It guarantees that consumer x does not imitate its own neighboring consumer on the left. In many classical models, the first-order IC condition is sufficient to guarantee incentive compatibility, so the first-order optimum is also the global optimum. However, in the bad news model here, the first-order optimal solution would violate the $R'(x)$ non-decreasing condition. So we need to introduce the second-order IC condition (the $R'(x)$ non-decreasing condition). More surprisingly, in some cases, the second-order IC condition does not guarantee full global incentive compatibility either, because consumers still have the incentive to double deviate.

Next, I will solve the bad news model in three steps in order to explain the impact of each IC condition. In the 5.1 section, I will solve the first-order optimum, that is, the relaxed problem additionally omitting the condition of non-decreasing $R'(x)$. In 5.2, I will solve the second-order optimum. And in 5.3, I will give the solution to the original problem which is immune to double deviation.

Now I rewrite the conditions and define a few things to facilitate the computation afterward. Condition (2) and (3) can be rewritten as

$$R'(x) \geq q(x) - \frac{G(\frac{q_1(x)}{Q})x[q(x) - q_1(x)]}{1 - (1 - x)G(\frac{q_1(x)}{Q})} \equiv J(x, q(x), q_1(x)),$$

and define

$$q_1^Q(x) = \arg\min_{q_1} J(x, Q, q_1), \tag{4}$$

and

$$K(x) = J(x, Q, q_1^Q(x)). \tag{5}$$

5.1 The First-order Optimum

Ignoring the condition that $R'(x)$ is non-decreasing, we have only the first-order IC condition (2), and we can obtain the first-order optimal solution in this section.

$$\max \int_0^1 [q(x)xf(x) - R'(x)(1 - F(x))]dx$$

$$s.t. \quad R'(x) \geq q(x) - \frac{G(\frac{q_1(x)}{Q})x[q(x) - q_1(x)]}{1 - (1-x)G(\frac{q_1(x)}{Q})} \equiv J(x, q(x), q_1(x)) \tag{6}$$

$$q(x) \in \{q_1(x), Q\}$$

After removing the $R'(x)$ non-decreasing condition, the problem becomes a pointwise problem. So condition (6) must bind. Note that condition (6) is obtained by rewriting the second stage IC condition. Thus in the first-order optimal solution, we will get a similar conclusion as in the good news model, i.e., the firm will not give any rent to the consumers in the second-stage upgrade, setting the upgrade price at a level where they are just willing to upgrade. This problem can be rewritten as

$$\max_{q_1, q \in \{q_1, Q\}} H(q_1, q, x) = qxf(x) - \left[q - \frac{xG(\frac{q_1}{Q})(q - q_1)}{1 - (1-x)G(\frac{q_1}{Q})}\right](1 - F(x)).$$

Note the function H is the integrand in the objective with $R'(x)$ being replaced by the lower bound J in equation (6).

The next lemma characterizes the shape of the function $H(\cdot, \cdot, x)$. It says for any type x, for any choice of q_1 and q, the integrand H is not as large as when we set both trial and upgrade size to 0, or when the upgrade size is full $q = Q$ and the trial size minimize the slope of the rent function. This allows us to show that a consumer either does not even get the trial, or his optimal upgrade would be full. Note, we are still agnostic about the trial size for those who will get a full upgrade.

Lemma 2. $H(q_1, q, x) \leq \max\{H(0, 0, x), H(q_1^Q(x), Q, x)\}$.

Proof. If

$$\frac{\partial H(q_1, q, x)}{\partial q} = xf(x) - \left[1 - \frac{xG(\frac{q_1}{Q})}{1 - (1-x)G(\frac{q_1}{Q})}\right](1 - F(x)) \leq 0,$$

then we have $H(q_1, q_1, x) \geq H(q_1, Q, x)$ and

$$xf(x) - (1 - F(x)) \leq -\frac{xG(\frac{q_1}{Q})}{1-(1-x)G(\frac{q_1}{Q})}(1 - F(x)) \leq 0.$$

Thus $H(0, 0, x) \geq H(q_1, q_1, x) \geq H(q_1, Q, x)$. Otherwise if

$$\frac{\partial H(q_1, q, x)}{\partial q} = xf(x) - \left[1 - \frac{xG(\frac{q_1}{Q})}{1-(1-x)G(\frac{q_1}{Q})}\right](1 - F(x)) \geq 0,$$

then we have $H(q_1, Q, x) \geq H(q_1, q_1, x)$. Because $q_1^Q(x)$ is the maximizer of $\max_{q_1} H(q_1, Q, x)$ (Recall the definition of $q_1^Q(x)$ in (4)), we further have $H(q_1^Q(x), Q, x) \geq H(q_1, Q, x) \geq H(q_1, q_1, x)$. □

So far, in the optimal schedule, we know a consumer is either not even given the trial, or will have a full upgrade. But we do not know who will get which. The following lemma tells us the structure of the schedule, if a type x gets a full upgrade, for any type x' larger than x, it's better also to give them a full upgrade than deny them even the trial. This leads us to the threshold structure in Proposition 3.

Lemma 3. *If $H(q_1^Q(x), Q, x) \geq H(0, 0, x)$, then for all $x' \geq x$, we have $H(q_1^Q(x'), Q, x') \geq H(0, 0, x')$.*

See Appendix C for proof.

Now I can formally state the optimal screening schedule for the first-order problem: there exists a threshold x_0 below which the consumers are denied trial (and, *a fortiori*, the upgrade); for the consumers whose types are above the threshold, their trial size is determined by minimizing the slope of the rent function and they always get a full upgrade if no bad news occurs.

Proposition 3. *In the bad news model, the first-order optimum is that: there exists x_0, such that for all $x \geq x_0$, $q(x) = Q$, $q_1(x) = q_1^Q(x)$, $R'(x) = K(x)$ and for all $x < x_0$, $q_1(x) = q(x) = 0$, and $R'(x) = 0$.*

Proof. By lemma 2, the first-order optimal $(q_1(x), q(x))$ is either $(0, 0)$ or $(q_1^Q(x), Q)$. Next, we just need to compare which of $(0, 0)$ and $(q_1^Q(x), Q)$ is better. When $x = 0$, $(0, 0)$ is better, and when $x = 1$, $(q_1^Q(x), Q)$ is better. Lemma 3 further tells us that once for an x, $(q_1^Q(x), Q)$ is better than $(0, 0)$, then for all $x' \geq x$, $(q_1^Q(x'), Q)$ is better than $(0, 0)$. So as x increases, once $q(x)$ has reached Q, it will stay at Q and will not go back to 0. □

By the proposition, since consumers with different x face different trial versions $q_1^Q(x)$, the first-order optimum is a rich menu.

Will this solution satisfy the second-order IC condition that $R'(x)$ is non-decreasing? By the envelope theorem,

$$R''(x) = K'(x) = \frac{\partial J(x, Q, q_1^Q(x))}{\partial x} \leq 0.$$

For the vast majority of functional forms, inequality is strict.

The next thing I am interested in is how $q_1^Q(x)$ will vary with x. In Courty and Li (2000), we see that what causes the first-order optimal solution not to satisfy the global incentive compatibility is that the quantity allocation does not increase with the prior and the posterior belief. In my model, this could be that for the posterior belief 0 (receiving the bad news), it is not the case that a larger x corresponds to a larger $q_1^Q(x)$.

Recall that $q_1^Q(x)$ is given by the following optimization.

$$\min_{q_1} J(x, Q, q_1) \quad \Leftrightarrow \quad \max_{q_1} \frac{G(\frac{q_1}{Q})(1 - \frac{q_1}{Q})}{1 - (1-x)G(\frac{q_1}{Q})}$$

Lemma 4. *In the first-order optimum, the trial size $q_1^Q(x)$ decreases in x.*

See Appendix D for proof.

Consistent with Courty and Li (2000), I also obtained in the first-order optimum that quantity allocation when receiving the bad news decreases in the prior type. This is one of the reasons that make the first-order optimum unable to satisfy the global incentive compatibility. This is intuitively natural because the a priori more optimistic consumers do not need a large trial version to learn the information.

5.2 The Second-order Optimum

The second-order problem is the first-order problem adding the $R'(x)$ non-decreasing condition. Unlike the first-order optimum which only considers local downward imitation, the second-order problem takes into account non-local imitation and upward imitation, only without double deviation. It is also the problem relaxed from the original problem by ignoring double deviation.

$$\max \int_0^1 [q(x)xf(x) - R'(x)(1 - F(x))]dx$$

$$s.t. \quad R'(x) \text{ is non-decreasing in } x \tag{7}$$

$$R'(x) = q(x) - p_2(x)G(\frac{q_1(x)}{Q})[q(x) - q_1(x)]$$

$$R'(x) \geq q(x) - \frac{G(\frac{q_1(x)}{Q})x[q(x) - q_1(x)]}{1 - (1-x)G(\frac{q_1(x)}{Q})} \equiv J(x, q(x), q_1(x)) \tag{8}$$

$$q(x) \in \{q_1(x), Q\}$$

Lemma 5. *In the second-order optimum of the bad news model, we have* $x_0 \leq x_1$ *such that in* $[0, x_0)$*,* $q(x) = q_1(x) = 0$*; in* $[x_0, x_1)$*,* $q(x) = q_1(x) \in (0, Q)$*; and in* $[x_1, 1]$*,* $q(x) = Q$*. If* $x_0 = x_1$*, then the second interval is an empty set.*

Proof. Firstly, I will prove that $q(x)$ increases in x. If $q(x)$ decreases in an interval $[x_1, x_2]$. The firm can get more revenue this way: cancel all plans on $(x_1, x_2]$ and let this group of consumers buy the plan for x_1. This gives us a flat $q(x)$, a flat $q_1(x)$ and a flat $R'(x)$ on $[x_1, x_2]$. This change does not violate constraint (8) because for all $x \geq x_1$,

$$R'(x) = R'(x_1) \geq J(x_1, q(x_1), q_1(x_1)) \geq J(x, q(x_1), q_1(x_1)),$$

where the last inequality is due to the negative partial derivative of J with respect to the first element. Recall that our objective value is $q(x)xf(x) - R'(x)(1 - F(x))$. With this change, revenue must improve because 1) $q(x)$ gets larger and 2) $R'(x)$ has to be non-decreasing thus a flat $R'(x)$ is no worse than before the change.

Because of the increasing property, $q(x)$ will not change back to 0 once it reaches a positive number, and once it reaches Q, it stays at Q and does not fall back again. □

By this lemma, we can divide the consumers into three parts, consumers of the low, medium, and high priors. High prior consumers will buy the trial version and will upgrade to the full version when they do not receive the bad news, medium prior consumers will buy only the trial version, and low prior consumers will not buy anything. And the two cut-off points are denoted by x_0 and x_1 respectively.

Define x^c as such that $\dfrac{x^c f(x^c)}{1 - F(x^c)} = 1$. This is where the marginal consumer is located when the firm sells an indivisible product. It is also the marginal consumer when the product is divisible but there is no learning in the trial.

In the following three lemmas, I show that all medium-prior consumers and all high-prior consumers are bunching respectively, and I also give out the relationship between the two cut-off points x_0, x_1 and the benchmark cut-off point x^c.

Lemma 6. *If $x_1 > x_0$, we have $x_0 = x^c$, and for all $x \in [x_0, x_1)$, $q(x) = q_1(x) = R'(x) = R'(x_1)$.*

Proof. If $x_1 > x_0$, in $[x_0, x_1)$, we have $q(x) = q_1(x) > 0$, $R'(x) = q_1(x)$, and the objective value is

$$q(x)xf(x) - R'(x)(1 - F(x)) = q_1(x)[xf(x) - (1 - F(x))] \tag{9}$$

If $x_0 < x^c$, it is better to have $q(x) = q_1(x) = 0$ in $[x_0, x^c)$. Thus we have $x_0 \geq x^c$. If $x_0 > x^c$, it is better to have $q_1(x) = q_1(x_0) > 0$ in $[x^c, x_0)$. So we should have $x_0 = x^c$. For all $x \in [x_0, x_1)$, to maximize the objective value in (9), we want $q_1(x) = R'(x)$ as high as possible, and the only constraint is that it is bounded by $R'(x_1)$ due to the non-decreasing $R'(x)$ condition. So to reach the maximum, we have $q_1(x) = R'(x) = R'(x_1)$. ◻

Lemma 7. *If $x_1 = x_0$, we have $x_1 = x_0 \leq x^c$.*

Proof. If $x_1 = x_0 > x^c$, it is better to have $q(x) = q_1(x) > 0$ in $[x^c, x_0)$. ◻

Lemma 8. *For all x in $[x_1, 1]$, $R'(x) = R'(x_1) = K(x_1)$, and $q_1(x) = q_1^Q(x_1)$.*

Proof. Either condition (7) or (8) binds. Condition (8) can be rewritten as

$$R'(x) \geq J(x, Q, q_1^Q(x)) = K(x)$$

Because of $\dfrac{\partial J(x, Q, q_1)}{\partial x} \leq 0$ and the envelope theorem, we have

$$K'(x) = \frac{\partial J(x, Q, q_1^Q(x))}{\partial x} \leq 0.$$

So in $[x_1, 1]$, condition (7) always binds and $R'(x) = R'(x_1)$. Next I will prove that $R'(x_1) = K(x_1)$, i.e. condition (8) will bind at x_1.

If $x_1 = x_0$, then for all $x \in [0, x_1)$, we have $R'(x) = 0$ because $q(x) = q_1(x) = 0$. Thus at x_1, condition (7) does not bind, condition (8) binds and $R'(x_1) = K(x_1)$.

If $x_1 > x_0$, by the previous lemma, we have $x_1 > x_0 = x^c$, and in $[x_0, x_1)$, we have $q(x) = q_1(x) = R'(x) = R'(x_1)$ and the objective value is $R'(x_1)[xf(x) - (1 - F(x))]$ where $xf(x) - (1 - F(x)) \geq 0$. If at x_1 condition (8) does not bind, i.e. $R'(x_1) > K(x_1)$, we can construct a better mechanism. For all x in the left neighborhood of x_1, i.e. for all $x \in (x_1 - \varepsilon, x_1)$, let $q(x) = Q$, $q_1(x) = q_1^Q(x)$ and $R'(x) = R'(x_1)$. This change does not violate any constraints because 1) a flat $R'(x)$ in $(x_1 - \varepsilon, x_1)$ is non-decreasing, 2) $K(x)$ is continuous in x, so that if $R'(x_1) > K(x_1)$, we should have $R'(x) = R'(x_1) \geq K(x)$ for all $x \in (x_1 - \varepsilon, x_1)$ if ε is small enough. So condition (8) is not violated either. Also, this change can bring in more revenue. The objective value after the change is

$Qxf(x) - R'(x_1)(1 - F(x))$, which is better than the objective value before the change. So we have $R'(x_1) = K(x_1) = \min_{q_1} J(x, Q, q_1) = J(x, Q, q_1^Q(x))$, which further gives us $q_1(x_1) = q_1^Q(x_1)$.

The solution $R'(x) = R'(x_1) = K(x_1)$ can be implemented by setting the same plan for everyone on $[x_1, 1]$, so we have $q_1(x) = q_1^Q(x_1)$. □

Now we can get to the optimal solution of the second-order problem.

Proposition 4. *In the bad news model, the second-order optimum without considering double deviation is,*

1. *with low priors $[0, x_0)$, consumers do not buy anything.*
2. *with medium priors $[x_0, x_1)$, consumers will buy the trial version, but never upgrade, even if they do not receive bad news. All the consumers in this interval share the same plan, i.e. the same trial size and the same prices of trial and upgrade. Their trial size is $K(x_1)$.*
3. *with high priors $[x_1, 1]$, consumers who do not receive bad news in the trial version will upgrade to the full product. All the consumers in this interval share the same plan, i.e. the same trial size and the same prices of trial and upgrade. Their trial size is $q_1^Q(x_1)$.*
4. *The medium-prior trial is larger than the high-prior trial.*

Proof. The first three points of this proposition can be directly obtained from the lemmas above.

$$K(x_1) = [Q - q_1^Q(x_1)] \frac{1 - G(\frac{q_1^Q(x_1)}{Q})}{1 - (1 - x)G(\frac{q_1^Q(x_1)}{Q})} + q_1^Q(x_1) \geq q_1^Q(x_1) \tag{10}$$

and the inequality is strict for most of the G functions. This means that consumers with medium priors, although they cannot upgrade, consume larger trial packs than those with high priors. □

Lemma 9. *A sufficient condition for $x_1 > x_0$ is*

$$\frac{x^c(1 - F(x^c))}{\int_{x^c}^1 (1 - F(x))dx} < \frac{1 - G(\frac{q_1^Q(x^c)}{Q})}{1 - (1 - x^c)G(\frac{q_1^Q(x^c)}{Q})}. \tag{11}$$

A necessary condition for $x_1 > x_0$ is

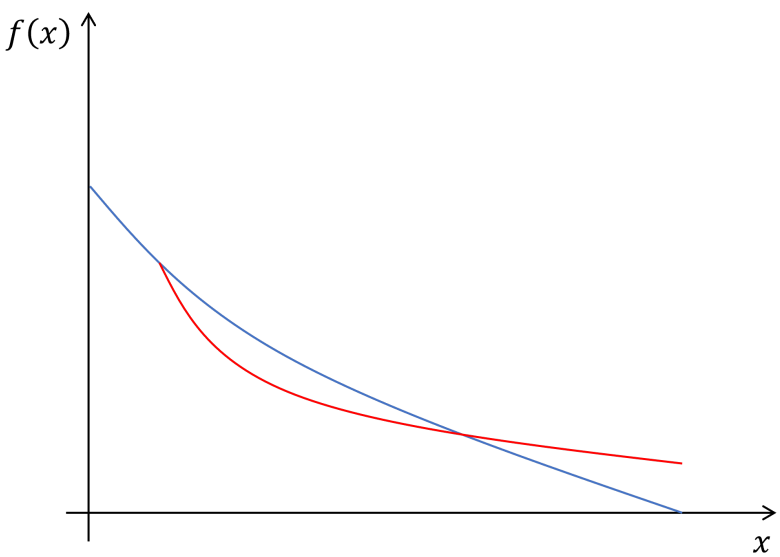
$$\frac{x^c(1 - F(x^c))}{\int_{x^c}^1 (1 - F(x))dx} < 1. \tag{12}$$

See Appendix E for proof.

When F and G satisfy the inequality (11), the second-order optimum is a two-part bunching. Otherwise if (12) is not met, it is a full bunching. In either case, the second-order IC condition gives us more bunching compared to the first-order optimal solution.

See examples of F and G that gives us $x_1 = x_0$ and $x_1 > x_0$ in Appendix F. Heavy-tailed F distributions are more likely to have $x_1 > x_0$, for example, truncated log-normal, truncated log-logistic, etc. More specifically, it is when the density curve $f(x)$ has a high density in the low x region, but the density does not decrease rapidly with increasing x, that is, there are also a sufficient number of consumers in the high x region. It is also easier to have condition (11) satisfied if there is a mass at $x = 1$.

In the figure below, the blue line is a typical probability density curve $f(x)$. If we keep the density at the low end unchanged, lower the density in the middle, and raise the density at the high end, as shown by the red line. The distribution shown by the red line would then satisfy condition (11).



When the F function has these properties, the firm faces a choice dilemma. It wants to charge high prices for high prior consumers but does not want to miss out on the large group of low prior consumers. So the best option for a firm is to offer a trial version to a large number of consumers, but an upgraded version only to a smaller number of consumers, i.e. a percentage of medium-prior consumers, even if they do not receive bad news, will not upgrade.

5.3 The Optimum with Full Incentive Compatibility

I have summarized that there are two cases of second-order optimal solutions, full bunching or two-part bunching. In the case of full bunching, we do not have double deviations because there is only one plan in the whole market. But in the case of two-part bunching, double deviation occurs. When F and G satisfy the inequality (11), in the second-order optimum, consumers on $[x_0, x_1)$ are incentivized to imitate x_1 and do not upgrade, even if they do not receive bad news.

Proposition 5. *In the bad news model, the second-order optimal solution obtained from the first- and second-order IC conditions does not satisfy full incentive compatibility when F and G satisfy the inequality* (11).

Proof. For a consumer x in $[x_0, x_1)$, if he buys his plan truthfully, he will get $R(x)$. Since the consumer x_1 has the second stage IC condition binding, all his rent comes from the first stage, i.e. $R(x_1) = q_1(x_1)x_1 - p_1(x_1)$. If consumer x buys the plan for x_1 but does not upgrade, he will get

$$\begin{aligned} q_1(x_1)x - p_1(x_1) &= R(x_1) - q_1(x_1)(x_1 - x) \\ &= R(x) + \int_x^{x_1} R'(t)dt - q_1(x_1)(x_1 - x) \\ &= R(x) + [K(x_1) - q_1^Q(x_1)](x_1 - x) \end{aligned}$$

where the last equality is because $R'(x) = K(x_1)$ for all $x \in [x_0, x_1)$ and $q_1(x_1) = q_1^Q(x_1)$. Because $K(x_1) \geq q_1^Q(x_1)$ by equation (10) and the inequality is strict with most of the functional forms, the consumer x in $[x_0, x_1)$ has an incentive to imitate x_1 and double deviate in the second stage. □

Therefore, with some functional forms of F and G, the second-order optimal solution is not our optimal solution either. Next, I will solve the optimal solution with full incentive compatibility.

Lemma 10. *In the optimal mechanism of the bad news model, we have $\tilde{x}_0 \leq \tilde{x}_1$ such that in $[0, \tilde{x}_0)$, $q(x) = q_1(x) = 0$; in $[\tilde{x}_0, \tilde{x}_1)$, $q(x) = q_1(x) \in (0, Q)$; and in $[\tilde{x}_1, 1]$, $q(x) = Q$. If $\tilde{x}_0 = \tilde{x}_1$, then the second interval is an empty set.*

Proof. I will prove the following two points.

1. If at some point x, $q(x) = Q$, then for all $x' > x$, we have $q(x') = Q$. We can construct

the optimal mechanism in $[x, 1]$ by letting

$$R'(x') = R'(x), \quad q(x') = q(x) = Q, \quad q_1(x') = q_1(x), \quad p_2(x') = p_2(x)$$

for all $x' > x$. Firstly, this mechanism is feasible. At x we have $R'(x) \geq J(x, Q, q_1(x))$. Because the partial derivative of J to the first element x is negative, we additionally have

$$R'(x') = R'(x) \geq J(x, Q, q_1(x)) \geq J(x', Q, q_1(x)) = J(x', Q, q_1(x'))$$

for all $x' > x$. And $R'(x)$ is flat in $[x, 1]$, so it is non-decreasing. Next, consider double deviation. This mechanism can be implemented by letting everyone on $[x, 1]$ buy the same plan. So if people on $[0, x)$ do not imitate x by double deviation, they will not imitate $(x, 1]$ by double deviation either. This mechanism does not violate any incentive compatibility conditions, then it is feasible. The objective value is $q(x)xf(x) - R'(x)(1 - F(x))$, and for all $x' > x$, the lower bound of $R'(x')$ is $R'(x)$ and the upper bound of $q(x')$ is Q. Thus there is no better mechanism. So once $q(x)$ reaches Q, it stays at Q and does not fall back again.

2. If at some point x, $q(x) = q_1(x) > 0$, then for all $x' > x$, we have $q(x') > 0$. At x, $R'(x) = q_1(x) > 0$. For all $x' > x$, if we have $q(x') = 0$, we also have $q_1(x') = 0$, then we get $R'(x) = 0$ violating the non-decreasing condition of $R'(x)$. So $q(x)$ will not change back to 0 once it reaches a positive number.

□

Similarly as in the second-order problem, we can divide the consumers into three parts, consumers of the low, medium, and high priors. High prior consumers will buy the trial version and will upgrade to the full version when they do not receive the bad news, medium prior consumers will buy only the trial version, and low prior consumers will not buy anything. And the two cut-off points are denoted by $\tilde{x}_0$ and $\tilde{x}_1$ respectively.

In the next lemma, I show that the high-prior consumers are bunching.

Lemma 11. *For all x in $[\tilde{x}_1, 1]$, $R'(x) = R'(\tilde{x}_1) = J(\tilde{x}_1, Q, q_1(\tilde{x}_1))$, and $q_1(x) = q_1(\tilde{x}_1)$.*

Proof. In the proof of lemma 10, I show that the optimal mechanism in $[\tilde{x}_1, 1]$ is such that for all $x > \tilde{x}_1$,

$$R'(x) = R'(\tilde{x}_1), \quad q(x) = q(\tilde{x}_1) = Q, \quad q_1(x) = q_1(\tilde{x}_1), \quad p_2(x) = p_2(\tilde{x}_1).$$

The next step is just to prove that $R'(\tilde{x}_1) = J(\tilde{x}_1, Q, q_1(\tilde{x}_1))$.

For all $x < \tilde{x}_1$, we have $q(x) = q_1(x) = R'(x)$ and the objective value is $R'(x)[xf(x) - (1 - F(x))]$. It is 0 if $R'(x) = 0$, otherwise it is positive and smaller than $R'(\tilde{x}_1)[xf(x) - (1 - F(x))]$ because $R'(\tilde{x}_1) \geq R'(x)$. If at $\tilde{x}_1$ we have $R'(\tilde{x}_1) > J(\tilde{x}_1, Q, q_1(\tilde{x}_1))$, we can construct a better mechanism with an objective value higher than $R'(\tilde{x}_1)[xf(x) - (1 - F(x))]$. For all x in the left neighborhood of $\tilde{x}_1$, i.e. for all $x \in (\tilde{x}_1 - \varepsilon, \tilde{x}_1)$, let $q(x) = Q$, $q_1(x) = q_1(\tilde{x}_1)$ and $R'(x) = R'(\tilde{x}_1)$. This change does not violate any constraints because 1) a flat $R'(x)$ in $(\tilde{x}_1 - \varepsilon, \tilde{x}_1)$ is non-decreasing, 2) $J(x, Q, q_1(\tilde{x}_1))$ is continuous in x, so that if $R'(\tilde{x}_1) > J(\tilde{x}_1, Q, q_1(\tilde{x}_1))$, we should have $R'(x) = R'(\tilde{x}_1) \geq J(x, Q, q_1(\tilde{x}_1))$ for all $x \in (x_1 - \varepsilon, x_1)$ if ε is small enough, 3) this change can be implemented by letting consumers in $(\tilde{x}_1 - \varepsilon, \tilde{x}_1)$ consume $\tilde{x}_1$'s plan, so if other consumers would not have imitated $\tilde{x}_1$ and double deviate before, they will not imitate others in $(\tilde{x}_1 - \varepsilon, \tilde{x}_1)$ and double deviate now after the change. The objective value after the change is $Qxf(x) - R'(\tilde{x}_1)(1 - F(x))$. This change can bring in more revenue.

□

In the next lemma, I give a simple and clear characterization of the mechanisms immune to double deviation, that is, the quantity of trial should be increasing in the prior among medium-prior consumers.

Lemma 12. *If $\tilde{x}_1 > \tilde{x}_0$, the condition of no double deviation is that $q_1(x)$ increases with x on $[\tilde{x}_0, \tilde{x}_1]$.*

Proof. By lemma 11, $R'(\tilde{x}_1) = J(\tilde{x}_1, Q, q_1(\tilde{x}_1))$, so consumer $\tilde{x}_1$ does not have any second-stage rent. If $\tilde{x}_1 > \tilde{x}_0$, in order for the consumers on $[\tilde{x}_0, \tilde{x}_1)$ not to imitate $\tilde{x}_1$ and double deviate, we need

$$R(\tilde{x}_1) - q_1(\tilde{x}_1)(\tilde{x}_1 - x) \leq R(x) \Leftrightarrow \frac{\int_x^{\tilde{x}_1} R'(x)dx}{\tilde{x}_1 - x} = \frac{\int_x^{\tilde{x}_1} q_1(x)dx}{\tilde{x}_1 - x} \leq q_1(\tilde{x}_1)$$

where we let x take $\tilde{x}_1 - \varepsilon$ and ε takes the limit to 0. Thus we get $\lim_{x \to \tilde{x}_1^-} q_1(x) \leq q_1(\tilde{x}_1)$. And since $R'(x) = q_1(x)$ must be increasing with x on $[\tilde{x}_0, \tilde{x}_1)$, we have the following condition: $q_1(x)$ increases with x on $[\tilde{x}_0, \tilde{x}_1]$. Easy to check that it is also sufficient. □

In the following two lemmas, I show some characterizations of the solution in the case of $\tilde{x}_1 > \tilde{x}_0$ and $\tilde{x}_1 = \tilde{x}_0$ respectively.

Lemma 13. *If $\tilde{x}_1 > \tilde{x}_0$, we have $\tilde{x}_0 = x^c$, and for all $x \in [\tilde{x}_0, \tilde{x}_1)$, $q(x) = q_1(x) = q_1(\tilde{x}_1)$.*

Proof. If $\tilde{x}_1 > \tilde{x}_0$, in $[\tilde{x}_0, \tilde{x}_1)$, we have $q(x) = q_1(x) > 0$, $R'(x) = q_1(x)$, and the objective value is

$$q(x)xf(x) - R'(x)(1 - F(x)) = q_1(x)[xf(x) - (1 - F(x))] \tag{13}$$

If $\tilde{x}_0 < x^c$, it is better to have $q(x) = q_1(x) = 0$ in $[\tilde{x}_0, x^c)$. Thus we have $\tilde{x}_0 \geq x^c$. If $\tilde{x}_0 > x^c$, it is better to have $q_1(x) = q_1(\tilde{x}_0) > 0$ in $[x^c, \tilde{x}_0)$. So we should have $\tilde{x}_0 = x^c$. For all $x \in [\tilde{x}_0, \tilde{x}_1)$, to maximize the objective value in (13), we want $q_1(x) = R'(x)$ as high as possible, and the constraint is that $q_1(x)$ increases with x on $[\tilde{x}_0, \tilde{x}_1]$. So to reach the maximum, we have $q_1(x) = q_1(\tilde{x}_1)$. □

Lemma 14. *If $\tilde{x}_1 = \tilde{x}_0$, we have $\tilde{x}_1 = \tilde{x}_0 \leq x^c$, $q_1(\tilde{x}_1) = q_1^Q(\tilde{x}_1)$ and $R'(\tilde{x}_1) = K(\tilde{x}_1)$.*

Proof. If $\tilde{x}_1 = \tilde{x}_0 > x^c$, it is better to have $q(x) = q_1(x) > 0$ in $[x^c, \tilde{x}_0)$, so $\tilde{x}_1 = \tilde{x}_0 \leq x^c$.

In $[\tilde{x}_1, 1]$, we maximize the objective value $Qxf(x) - J(\tilde{x}_1, Q, q_1(\tilde{x}_1))(1 - F(x))$ under the following two constraints, 1) $R'(\tilde{x}_1) = J(\tilde{x}_1, Q, q_1(\tilde{x}_1)) \geq R'(x)$, for all $x < \tilde{x}_1$, and 2) $q_1(\tilde{x}_1) \geq q_1(x)$, for all $x < \tilde{x}_1$. If $\tilde{x}_1 = \tilde{x}_0$, then for all $x \in [0, \tilde{x}_1)$, we have $R'(x) = 0$ because $q(x) = q_1(x) = 0$. Thus at $\tilde{x}_1$, the first constraint does not bind. Also because $q_1(x) = 0$ for all $x < \tilde{x}_1$, the second constraint does not bind either. So we should choose $q_1(\tilde{x}_1) = q_1^Q(\tilde{x}_1)$ to minimize $J(\tilde{x}_1, Q, q_1(\tilde{x}_1))$ and thus $R'(\tilde{x}_1) = J(\tilde{x}_1, Q, q_1^Q(\tilde{x}_1)) = K(\tilde{x}_1)$. □

Now we are able to get to the final optimal mechanism, which is a full bunching.

Proposition 6. *In the bad news model, the optimum with full incentive compatibility can be implemented by offering only one plan, i.e. one trial size, one trial price and one upgrade unit price. High-prior consumers in $[\tilde{x}_1, 1]$ will upgrade to the full product if they do not receive bad news. Medium-prior consumers in $[\tilde{x}_0, \tilde{x}_1)$ will only consume the trial. And low-prior consumers in $[0, \tilde{x}_0)$ do not buy anything.*

Proof. Note that the size of the trial version purchased by consumers of $[\tilde{x}_0, \tilde{x}_1)$ and consumers of $[\tilde{x}_1, 1]$ are both $q_1(\tilde{x}_1)$. So the optimum with full incentive compatibility can be implemented by full bunching, i.e. offering only one plan. This is because we set the upgrade unit price just so that consumer $\tilde{x}_1$ will buy it (by lemma 11, $R'(\tilde{x}_1) = J(\tilde{x}_1, Q, q_1(\tilde{x}_1))$, thus no second-stage rent for $\tilde{x}_1$), consumers on $[\tilde{x}_0, \tilde{x}_1)$ will naturally not upgrade, even if they do not receive bad news. □

Lemma 15. *A sufficient condition for $\tilde{x}_1 > \tilde{x}_0$ is*

$$\frac{x^c(1 - F(x^c))}{\int_{x^c}^1 (1 - F(x))dx} < \frac{1 - G(\frac{q_1^Q(x^c)}{Q})}{1 - (1 - x^c)G(\frac{q_1^Q(x^c)}{Q})}, \tag{14}$$

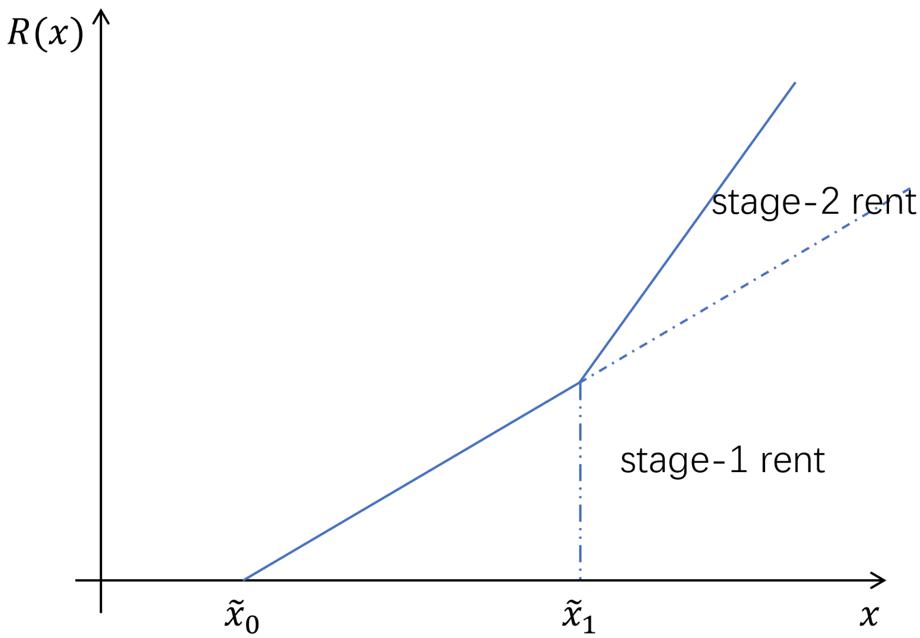
A necessary condition for $\tilde{x}_1 > \tilde{x}_0$ is

$$\frac{x^c(1 - F(x^c))}{\int_{x^c}^1 (1 - F(x))dx} < 1. \tag{15}$$

Proof. Condition (15) is naturally necessary. If it does not hold, we have $x_1 = x_0$, and get full bunching in the second-order optimum, which is immune to double deviation. Thus we have $x_1 = x_0 = \tilde{x}_1 = \tilde{x}_0$. I can also show that (14) is sufficient by exactly the same steps as I did in Appendix E to get a contradiction to lemma 14. □

Here the conditions of $\tilde{x}_1 > \tilde{x}_0$ are the same as the previous conditions of $x_1 > x_0$. When there are very many low-prior consumers, but also a sufficient number of high-prior consumers, the firm sells the trial version to more consumers, but sets the price of the upgrade higher so that a fraction of middle-prior consumers will only buy the trial version even if no bad news occur.

The consumer's rent curve is shown in the figure. Although the optimal solution is full bunching, the rent curve is not straight. It has a kink. This is because, starting from $\tilde{x}_1$, a new part of the rent brought by the second stage is added.



5.4 Double Deviation and Rent Allocation Between Stages

Why exactly is there double deviation? It is because of a certain asymmetry: the condition that consumers do not imitate to the left is not the same as the condition that they do not imitate to the right. While we compute the IC condition from downward imitation, the condition for upward imitation is ignored. Because consumers are not guaranteed not to get negative rents even in the worst-case scenario, when consumers imitate consumers who are more optimistic than they are, they are likely not to upgrade because that would give them negative rents even if they do not receive bad news.

Another way to understand the issue is to think about the degrees of freedom in this problem: to pin down the firm's mechanism, how many dimensions does it have the freedom to choose? In standard mechanism design problems, just by determining $R(x)$ we have determined everything. But here $R(x)$ is not enough. It is equally important whether $R(x)$ is placed in the first or second stage, which on the one hand affects the revenue as we have seen in section 3, and on the other hand, affects the incentive compatibility through asymmetric imitation and double deviation.

Denote by $R_2(x)$ the second stage rent of consumer x. In order for x not to mimic $x' > x$ and never upgrade in the second stage, we need to satisfy the following condition.

$$R(x) \geq R(x') - R_2(x') - q_1(x')(x' - x)$$

where $R(x') - R_2(x')$ is the first-stage rent of consumer x'. The rent of the consumer x imitating x' and not upgrading is the first-stage rent of x' minus the utility bonus of consumer x' of consuming the same trial version $q_1(x')$ over x. This gives us a lower bound on $R_2(x)$, which is also an upper bound on $p_2(x)$. To avoid double deviation, we need to leave enough rent in the second stage. In the models in the above sections, we can get $\tilde{x}_1 < x_1$ relative to most functional forms. Further, we can get that the upgrade price in the optimal solution is cheaper than the upgrade price in the second-order optimal solution. In other words, to avoid double deviation, the firm adjusts upward the allocation of the second-order rent.

So in summary, in addition to deciding on each consumer's rent, we also need to decide how the rent will be distributed between the two stages, in order to maximize revenue and also to meet incentive compatibility.

6 Belief Updating and Bunching

When consumers' priors are more homogeneous, firms are always better able to squeeze the consumer surplus and make maximum profits. So when consumers' priors are heterogeneous, companies always find ways to minimize the rent difference. In this two-stage problem, they will try to raise the upgrade price as much as possible, so as to reduce the advantage of optimistic consumers over pessimistic consumers. Because even if the optimistic consumer is in the good scenario with a higher probability, they still need to pay a high price to get the upgrade, which will somewhat wipe out his a priori advantage. This is what a first-order optimal contract looks like.

But would a first-order optimal contract be truly optimal? In the bad news model, it is not. In the bad news model, the signals are not positive enough, and even for consumers who are in the good scenario, their willingness to pay for the goods still depends heavily on their prior knowledge. This results in the fact that conditional on not receiving bad news, a high-prior consumer will still be more willing to pay for the upgrade than a low-prior consumer. On the other hand, in order to raise the upgrade price as much as possible, the firm will leave consumers no rent in the second stage, which means pricing the upgrade just to everyone's willingness to pay. The firm should give optimistic consumers a more expensive upgrade and a relatively cheaper trial, but this is exactly what the most pessimistic consumers want because they are less concerned about the upgrade price and more concerned about the trial. Meanwhile, the firm will give pessimistic consumers a relatively cheap upgrade because their willingness to pay is lower, and this is exactly what the most optimistic consumers want because they have a high probability of upgrading. So this causes a first-order optimal contract not to hold, and it collapses by non-local imitation. This kind of imitation makes it costly for firms to fully discriminate without violating the second-order IC condition, so it bunches more in the second-order optimum. Double deviation makes this type of imitation more likely to happen because low-prior consumers can buy trial versions of high-prior consumers, but not upgrade even without any bad news. So after considering double deviation, we get full bunching.

But in the good news model, the situation is different because good news can erase the a priori differences. So there is no such non-local imitation of each other.

The conclusion we can draw from here is that whether a firm has the incentive to offer a rich menu depends on whether its trial version can give sufficiently positive signals to erase the differences in priors to a certain extent.

7 Application and Discussion

The model can always be considered in the context of bank lending with heterogeneous entrepreneurs, each privately informed about the prospect of their projects, i.e., the ex-ante type of the project. Whether the project will be successful requires an initial investment and the entrepreneur could privately learn the probability that the project will yield a positive cash flow. The bank could lend to the firm sequentially: the firm obtains partial funding for the initial investment and can borrow additional fundings depending on the signal generated from the initial investment. The bank screens the firms through the quantity of the initial investment as well as the price of the capital, a.k.a., interest rates in both stages. This type of contract can be implemented in various forms, for example, an initial loan with a contingent line of credit; relationship lending; etc. Sequential lending has been studied extensively in finance, for example, Bizer and DeMarzo (1992), but little attention is paid to sequential screening in lending relationships[9]. Echoing my main results on bunching, Cox et al. (2021) empirically shows that bunching is a salient feature of small business lending.

This model can also be applied to the labor market, as internships can be interpreted as a kind of trial version. Policymakers, as a principal, can design different mechanisms through various internship and wage system policies. The company is the agent and the employee is the equivalent of the commodity in my model. I maximize social surplus minus the agent's rent, which in the labor market can be interpreted as social surplus minus the firm's rent, i.e. workers' welfare. Bardhi et al. (2020) divides different jobs into "breakthrough" jobs (star jobs) and "breakdown" jobs (guardian jobs), which is consistent with my characterization of good news and bad news belief updating.

8 Concluding Remarks

Different information structures led sellers to adopt different trial options. In good news belief updates, like video games, companies have the incentive to provide rich trial menus, while in bad news belief updates, like mobile plans, companies only want to provide a single trial version.

With good news belief updating, the firm has set the price of the upgrade very high, leaving the consumer with no rent at all in the second stage. All the rents are offered in the first stage trial version. More optimistic consumers will consume larger samples. On the one hand, the unit price of samples is cheaper than the unit price of upgrades, so optimistic consumers are willing to consume a larger sample; on the other hand, companies are willing to offer a larger trial to optimistic consumers because a larger trial is more likely to bring them the good news and raise their willingness to pay.

With bad news belief updating, the optimal mechanism can be implemented by offering only one plan, i.e. one trial size, one trial price, and one upgrade unit price. High-prior consumers will upgrade to the full product if they do not receive bad news, and they do have second-stage rent. Low-prior consumers do not buy anything. There are also medium-prior consumers if the type distribution is heavy-tailed. Those medium-prior consumers will only consume the trial, and never upgrade even if there is no bad news.

[9] Chowdhury (2007) also studies sequential financing and contingent renewal but sequential financing is only a partial screening mechanism in his model.

References

Bardhi, A., Y. Guo, and B. Strulovici (2020). Early-career discrimination: Spiraling or self-correcting? Technical report, Working Paper, Duke and Northwestern.

Baron, D. P. and D. Besanko (1984). Regulation and information in a continuing relationship. *Information Economics and policy 1*(3), 267–302.

Battaglini, M. (2005). Long-term contracting with markovian consumers. *American Economic Review 95*(3), 637–658.

Battaglini, M. and R. Lamba (2019). Optimal dynamic contracting: The first-order approach and beyond. *Theoretical Economics 14*(4), 1435–1482.

Bergemann, D. and S. Morris (2019). Information design: A unified perspective. *Journal of Economic Literature 57*(1), 44–95.

Bergemann, D. and J. Välimäki (2006). Dynamic pricing of new experience goods. *Journal of Political Economy 114*(4), 713–743.

Bergemann, D. and J. Välimäki (2019). Dynamic mechanism design: An introduction. *Journal of Economic Literature 57*(2), 235–74.

Bhargava, H. K. and V. Choudhary (2008). Research note—when is versioning optimal for information goods? *Management Science 54*(5), 1029–1035.

Bizer, D. S. and P. M. DeMarzo (1992). Sequential banking. *Journal of Political Economy 100*(1), 41–61.

Boleslavsky, R. and M. Said (2013). Progressive screening: Long-term contracting with a privately known stochastic process. *Review of Economic Studies 80*(1), 1–34.

Börgers, T. and D. Krahmer (2015). *An introduction to the theory of mechanism design*. Oxford University Press, USA.

Chellappa, R. K. and A. Mehra (2018). Cost drivers of versioning: Pricing and product line strategies for information goods. *Management Science 64*(5), 2164–2180.

Chellappa, R. K. and S. Shivendu (2005). Managing piracy: Pricing and sampling strategies for digital experience goods in vertically segmented markets. *Information Systems Research 16*(4), 400–417.

Chowdhury, P. R. (2007). Group-lending with sequential financing, contingent renewal and social capital. *Journal of development Economics 84*(1), 487–506.

Courty, P. and H. Li (2000). Sequential screening. *The Review of Economic Studies 67*(4), 697–717.

Cox, N., E. Liu, D. Morrison, et al. (2021). Market power in small business lending: A two-dimensional bunching approach. Technical report.

Eső, P. and B. Szentes (2007a). Optimal information disclosure in auctions and the handicap auction. *The Review of Economic Studies 74*(3), 705–731.

Eső, P. and B. Szentes (2007b). The price of advice. *The Rand Journal of Economics 38*(4), 863–880.

Gentzkow, M. and E. Kamenica (2014). Costly persuasion. *American Economic Review 104*(5), 457–62.

Guo, H., W. He, and B. Liu (2022). Learning by consuming: Optimal pricing with endogenous information provision. *arXiv preprint arXiv:2209.01453*.

Kamenica, E. (2019). Bayesian persuasion and information design. *Annual Review of Economics 11*, 249–272.

Kamenica, E. and M. Gentzkow (2011). Bayesian persuasion. *American Economic Review 101*(6), 2590–2615.

Li, H. and X. Shi (2017). Discriminatory information disclosure. *American Economic Review 107*(11), 3363–85.

Matyskova, L. (2018). Bayesian persuasion with costly information acquisition. *cERGE-EI Working Paper Series* (614).

Nguyen, A. and T. Y. Tan (2021). Bayesian persuasion with costly messages. *Journal of Economic Theory 193*, 105212.

Pavan, A., I. Segal, and J. Toikka (2014). Dynamic mechanism design: A myersonian approach. *Econometrica 82*(2), 601–653.

Wei, D. and B. Green (2019). Price discrimination with information design. *Available at SSRN*.

A Proof of lemma 1

Proof. Define $r(x, x')$ as the rent where the consumer x imitates x' and does not double deviate in the second stage.

$$r(x, x') = [q_1(x')x - p_1(x')] + xG(\frac{q_1(x')}{Q})[q(x') - q_1(x')][1 - p_2(x')]$$

in the good news model, and

$$r(x, x') = q(x')x - p_1(x') - p_2(x')[q(x') - q_1(x')]\left[1 - (1 - x)G(\frac{q(x')}{Q})\right]$$

in the bad news model.

First, show they are necessary. By the first-stage IC condition,

$$R(x) = \max_{x'} r(x, x'),$$

and by the envelope theorem,

$$R'(x) = \left.\frac{\partial r(x, x')}{\partial x}\right|_{x' \to x} = q_1(x) + G(\frac{q_1(x)}{Q})[q(x) - q_1(x)][1 - p_2(x)]$$

in the good news model, and

$$R'(x) = \left.\frac{\partial r(x, x')}{\partial x}\right|_{x' \to x} = q(x) - p_2(x)G(\frac{q_1(x)}{Q})[q(x) - q_1(x)]$$

in the bad news model.

Moreover, for all $x' < x$,

$$R(x) - R(x') \geq r(x, x') - r(x', x') = (x - x')R'(x')$$

$$R(x) - R(x') \leq r(x, x) - r'(x', x) = (x - x')R'(x)$$

Thus,

$$R'(x') \leq \frac{R(x) - R(x')}{x - x'} \leq R'(x),$$

so $R'(x)$ weakly increases in x.

Next, show they are sufficient in the relaxed problem. For any x, x',

$$\begin{aligned} r(x,x) - r(x,x') &= [r(x,x) - r(x',x')] - [r(x,x') - r(x',x')] \\ &= [R(x) - R(x')] - [r(x,x') - r(x',x')] \\ &= \int_{x'}^{x} R'(s)ds - (x - x')R'(x') \\ &= \int_{x'}^{x} [R'(s) - R'(x')]ds \geq 0 \end{aligned}$$

where the last inequality uses $R'(x)$ weakly increasing in x. So we have that for all x, x',

$$r(x,x) \geq r(x,x'),$$

and the first-stage IC condition in the relaxed problem is met. □

B Proof of $q_1(x)$ increasing in x in the good news model

$$\begin{aligned} \max_{q_1} V(q_1,x) &\Leftrightarrow \max_{q_1} \frac{V(q_1,x)}{xf(x)} \\ &\Leftrightarrow \max_{q_1} M(q_1|x) = N(q_1) - q_1\frac{1-F(x)}{xf(x)} \end{aligned}$$

where

$$N(q_1) = q_1 + G(\frac{q_1}{Q})(Q - q_1).$$

Write $q_1^\star(x)$ as the optimal q_1 maximizing $M(q_1|x)$. I will show that for all $x' \geq x$, we have $q_1^\star(x') \geq q_1^\star(x)$.

For all q_1,

$$N(q_1^\star(x)) - q_1^\star(x)\frac{1-F(x)}{xf(x)} \geq N(q_1) - q_1\frac{1-F(x)}{xf(x)}.$$

Thus for all $q_1 \leq q_1^\star(x)$, and for all $x' \geq x$,

$$N(q_1^\star(x)) - N(q_1) \geq [q_1^\star(x) - q_1]\frac{1-F(x)}{xf(x)} \geq [q_1^\star(x) - q_1]\frac{1-F(x')}{x'f(x')}$$

$$\Rightarrow N(q_1^\star(x)) - q_1^\star(x)\frac{1-F(x')}{x'f(x')} \geq N(q_1) - q_1\frac{1-F(x')}{x'f(x')}.$$

So that for all $q_1 \leq q_1^\star(x)$, $M(q_1^\star(x)|x') \geq M(q_1|x')$, which further gives us $q_1^\star(x') \geq q_1^\star(x)$.

C Proof of lemma 3

$$H(q_1, Q, x) = Q\left[xf(x) - \left(1 - \frac{G(\frac{q_1}{Q})x(1-\frac{q_1}{Q})}{1-(1-x)G(\frac{q_1}{Q})}\right)(1-F(x))\right]$$

$$H(q_1, Q, x) \geq 0 \quad \Leftrightarrow \quad P(q_1, x) \leq \frac{xf(x)}{1-F(x)} \tag{16}$$

where

$$P(q_1, x) = \frac{1-(1-\frac{q_1}{Q}x)G(\frac{q_1}{Q})}{1-(1-x)G(\frac{q_1}{Q})}.$$

In (16), the right-hand-side term increases in x by assumption (1), and the left-hand-side term decreases in x because

$$\frac{\partial P(q_1, x)}{\partial x} = -\frac{G(\frac{q_1}{Q})[1-G(\frac{q_1}{Q})](1-\frac{q_1}{Q})}{[1-(1-x)G(\frac{q_1}{Q})]^2} \leq 0.$$

So

$$H(q_1^Q(x), Q, x) \geq 0 = H(0,0,x) \Leftrightarrow P(q_1^Q(x), x) \leq \frac{xf(x)}{1-F(x)}.$$

For all $x' \geq x$,

$$P(q_1^Q(x), x') \leq P(q_1^Q(x), x) \leq \frac{xf(x)}{1-F(x)} \leq \frac{x'f(x')}{1-F(x')} \Rightarrow H(q_1^Q(x), Q, x') \geq 0.$$

Because $q_1^Q(x') = \arg\max_{q_1} H(q_1, Q, x')$, we have

$$H(q_1^Q(x'), Q, x') \geq H(q_1^Q(x), Q, x') \geq 0.$$

D Proof of lemma 4

By assumption 2, the elasticity of $M(\frac{q_2}{Q})$ increases in q_2. Define $z_1 = \frac{q_1}{Q}$ and $z_2 = \frac{q_2}{Q}$. Recall that $M(z_2) = 1 - G(z_1) = 1 - G(1-z_2)$. Thus we have

$$\frac{dM}{dq_2}\Big/\frac{M}{q_2} = \frac{dM}{dz_2}\Big/\frac{M}{z_2} = \frac{g(1-z_2)z_2}{1-G(1-z_2)} = \frac{g(z_1)(1-z_1)}{1-G(z_1)} \tag{17}$$

decreasing in z_1. Write $G^{-1}(b)$ as the inverse function of $G(z_1)$, then the optimization with respect to z_1 can be rewritten as

$$\max_b y(b) = \frac{b(1-G^{-1}(b))}{1-(1-x)b}$$

$$\Leftrightarrow \max_b \log y(b) = \log(b) + \log(1-G^{-1}(b)) - \log(1-(1-x)b).$$

By (17) and $g(z_1)(G^{-1})'(b) = 1$, we get that

$$\frac{1-G(z_1)}{g(z_1)(1-z_1)} = (1-b)\frac{(G^{-1})'(b)}{1-G^{-1}(b)} \tag{18}$$

increases in b.

$$[\log y(b)]' = \frac{1}{b} - \frac{(G^{-1})'(b)}{1-G^{-1}(b)} + \frac{1-x}{1-(1-x)b} \tag{19}$$

$$[\log y(b)]' \geq 0 \Leftrightarrow b[1-(1-x)b]\frac{(G^{-1})'(b)}{1-G^{-1}(b)} \leq 0$$

$$\Leftrightarrow b[1-(1-x)b]\frac{(G^{-1})'(b)}{1-G^{-1}(b)} = b(1-b)\frac{(G^{-1})'(b)}{1-G^{-1}(b)} \cdot \frac{1-(1-x)b}{1-b} \leq 0 \tag{20}$$

We know from (18) that $(1-b)\frac{(G^{-1})'(b)}{1-G^{-1}(b)}$ increases in b. And easy to check that $\frac{1-(1-x)b}{1-b}$ also increases in b. Thus (20) increases in b. So the first-order condition $[\log y(b)]' = 0$ is sufficient for a global optimum.

By (19), as x grows, $[\log y(b)]'$ decreases, so the optimal b will move to the left. Since $b = G(\frac{q_1^Q(x)}{Q})$, $q_1^Q(x)$ also decreases in x.

E Proof of lemma 9

Firstly, I show that (11) is sufficient, that is, whenever the inequality (11) holds, we must have $x_1 > x_0$. I prove this statement by contradiction. Suppose the inequality (11) holds but $x_1 = x_0$ in the optimal mechanism. In what follows, I will find the optimal threshold x_1 within the class of mechanisms satisfying $x_0 = x_1$ and show that it must be larger than x^c, thus contradicting Lemma 7.

Step 1 I first characterize the optimality condition of x_1. Given $x_1 = x_0$, to find the optimal

x_1, I solve the following problem.

$$\max_{x_1} V(x_1) = \int_{x_1}^{1} [Qxf(x) - K(x_1)(1 - F(x))]dx \tag{21}$$

Using Leibniz's rule, I take the derivative of $V(x_1)$ with respect to x_1

$$V'(x_1) = -K'(x_1)\int_{x_1}^{1}(1 - F(x))dx - Qx_1 f(x_1) + K(x_1)(1 - F(x_1))$$

Step 2 I show Equation (11) implies $V'(x^c) > 0$. Evaluating $V'(x_1)$ at $x_1 = x^c$ and using $x^c f(x^c) = 1 - F(x^c)$, I obtain

$$V'(x^c) = -K'(x^c)\int_{x^c}^{1}(1 - F(x))dx - (Q - K(x^c))(1 - F(x^c))$$

Recall from the definition of $K(x)$ in Equation (5) and taking its derivative w.r.t. x

$$K(x^c) = Q - \frac{G(\frac{q_1^Q(x^c)}{Q})x^c(Q - q_1^Q(x^c))}{1 - (1 - x^c)G(\frac{q_1^Q(x^c)}{Q})} \text{ and } K'(x^c) = \frac{(1 - G(\frac{q_1^Q(x^c)}{Q}))G(\frac{q_1^Q(x^c)}{Q})(Q - q_1^Q(x^c))}{[1 - (1 - x^c)G(\frac{q_1^Q(x^c)}{Q})]^2}$$

The condition $V'(x^c) > 0$ is equivalent to Equation (11) when I substitute $K(x^c)$ and $K'(x^c)$ in, and simplify it.

Step 3 I will show for all $x \le x^c$, $V'(x) > 0$. The condition $V'(x) > 0$ is equivalent to

$$-K'(x)\int_{x}^{1}(1 - F(x))dx > (Q - K(x))(1 - F(x)) - Q[(1 - F(x)) - xf(x)]$$

For all $x \le x^c$, the last term $(1 - F(x)) - xf(x) \ge 0$, so a sufficient condition for $V'(x) > 0$ is that

$$-K'(x)\int_{x}^{1}(1 - F(x))dx > (Q - K(x))(1 - F(x)),$$

which, if I substitue $K(x)$ and $K'(x)$ in, is equivalent to

$$\frac{x(1 - F(x))}{\int_x^1 (1 - F(t))dt} < \frac{1 - G(\frac{q_1^Q(x)}{Q})}{1 - (1 - x)G(\frac{q_1^Q(x)}{Q})}. \tag{22}$$

It is easy to see that the left-hand-side term increases with x in $[0, x^c]$ as

$$\frac{\mathrm{d}}{\mathrm{d}x}\frac{x(1-F(x))}{\int_x^1(1-F(t))dt} = \frac{\int_x^1(1-F(t))dt[1-F(x)-xf(x)]+x(1-F(x))^2}{\left(\int_x^1(1-F(t))dt\right)^2} > 0, \forall x \in [0, x^c]. \tag{23}$$

And we know that the condition (22) is satisfied at x^c by inequality (11). It suffices to show that the RHS is decreasing in x on $[0, x^c]$.

Define $\tilde{G}(x) = G(\frac{q_1^Q(x)}{Q})$ to simplify the notations. Taking the derivative of the RHS of inequality (22), we find the RHS decreases in x if and only if

$$-\tilde{G}'(x)[1-(1-x)\tilde{G}(x)] - (1-\tilde{G}(x))[\tilde{G}(x) - (1-x)\tilde{G}'(x)] \le 0$$

$$\Leftrightarrow -x\tilde{G}'(x) - (1-\tilde{G}(x))\tilde{G}(x) \le 0 \tag{24}$$

$$\Leftrightarrow -\frac{\tilde{G}'(x)}{\tilde{G}(x)/x} \le 1-\tilde{G}(x)$$

$$\Leftrightarrow -\frac{dG(\frac{q_1^Q(x)}{Q})}{dx}\Bigg/\frac{G(\frac{q_1^Q(x)}{Q})}{x} \le 1 - G(\frac{q_1^Q(x)}{Q}). \tag{25}$$

In the proof of lemma 4, I get that $q_1^Q(x)$ is solved by $[\log y(b)]' = 0$, where $b = G(\frac{q_1^Q(x)}{Q})$.

$$[\log y(b)]' = 0 \Leftrightarrow x = 1 - \frac{A(b)}{1+bA(b)}$$

where

$$A(b) = \frac{(G^{-1})'(b)}{1-G^{-1}(b)} - \frac{1}{b}. \tag{26}$$

$$\frac{dx}{db} = -\frac{A'(b) - A^2(b)}{(1+A(b)b)^2}$$

$$-\frac{dx}{db}\Bigg/\frac{x}{b} = \frac{bA'(b) - bA^2(b)}{(1+A(b)b)^2 - A(b) - A^2(b)b} \tag{27}$$

By (27), I get

$$(25) \Leftrightarrow -\frac{dG(\frac{q_1^Q(x)}{Q})}{dx}\Bigg/\frac{G(\frac{q_1^Q(x)}{Q})}{x} = \frac{(1+A(b)b)^2 - A(b) - A^2(b)b}{bA'(b) - bA^2(b)} \le 1-b \tag{28}$$

Define

$$m(b) = \frac{(G^{-1})'(b)}{1 - G^{-1}(b)}.$$

Thus,

$$(28) \Leftrightarrow [bm(b)]^2 - m(b) + \frac{1}{b} - b\left[m(b) - \frac{1}{b}\right]^2 \leq b(1-b)\left[m'(b) + \frac{1}{b^2} - \left(m(b) - \frac{1}{b}\right)^2\right]$$

$$\Leftrightarrow m'(b)b(1-b) + m(b)(1-2b) \geq 0$$

$$\Leftrightarrow [b(1-b)m(b)]' \geq 0$$

Under assumption 2, I show that $(1-b)m(b)$ increases in b in the proof of lemma 4, equation (18). Thus I have $b(1-b)m(b)$ increasing in b.

Step 4 So far I proved $V'(x) > 0$ for all $x \in [0, x^c]$, therefore, the optimal threshold x_1 must be larger than x^c, which contradicts Lemma 7.

Step 5 We cannot have $x_1 = x_0$ when (11) holds. Thus (11) leads to $x_1 > x_0$.

Secondly, show that (12) is necessary. Again I prove this statement by contradiction. Suppose the inequality (12) does not hold but $x_1 > x_0$. In what follows, I will find the optimal threshold x_1 within the class of mechanisms satisfying $x_1 > x_0$ and show that all $x_1 \in (x_0, 1]$ is dominated by $x_1 = x_0$.

Step 1 I first characterize the optimality condition of x_1. Given $x_1 > x_0$, by Lemma 6, we have $x_0 = x^c$. To find the optimal x_1, I solve the following problem.

$$\max_{x_1 \in (x^c, 1]} W(x_1) = \int_{x^c}^{x_1} K(x_1)[xf(x) - (1 - F(x))]dx + \int_{x_1}^{1} [Qxf(x) - K(x_1)(1 - F(x))]dx$$

Using Leibniz's rule, I take the derivative of $W(x_1)$ with respect to x_1,

$$W'(x_1) = K'(x_1)\left[\int_{x^c}^{x_1} xf(x)dx - \int_{x^c}^{1}(1 - F(x))dx\right] - (Q - K(x_1))x_1 f(x_1)$$

Step 2 I show that if (12) does not hold, we have $W'(x^c) \leq 0$. Evaluating $W'(x_1)$ at $x_1 = x^c$ and using $x^c f(x^c) = 1 - F(x^c)$, I obtain

$$W'(x^c) = -K'(x^c)\int_{x^c}^{1}(1 - F(x))dx - (Q - K(x^c))(1 - F(x^c))$$

Recall again from the definition of $K(x)$ in Equation (5) and taking its derivative w.r.t. x,

$$K(x^c) = Q - \frac{G(\frac{q_1^Q(x^c)}{Q})x^c(Q - q_1^Q(x^c))}{1-(1-x^c)G(\frac{q_1^Q(x^c)}{Q})} \text{ and } K'(x^c) = \frac{(1-G(\frac{q_1^Q(x^c)}{Q}))G(\frac{q_1^Q(x^c)}{Q})(Q - q_1^Q(x^c))}{[1-(1-x^c)G(\frac{q_1^Q(x^c)}{Q})]^2}.$$

I substitute $K(x^c)$ and $K'(x^c)$ in, and simplify it, then the condition $W'(x^c) \leq 0$ is equivalent to

$$\frac{x^c(1-F(x^c))}{\int_{x^c}^1 (1-F(x))dx} \geq \frac{1-G(\frac{q_1^Q(x^c)}{Q})}{1-(1-x^c)G(\frac{q_1^Q(x^c)}{Q})}.$$

If the inequality (12) does not hold, I have $W'(x^c) \leq 0$ because

$$\frac{x^c(1-F(x^c))}{\int_{x^c}^1 (1-F(x))dx} \geq 1 \geq \frac{1-G(\frac{q_1^Q(x^c)}{Q})}{1-(1-x^c)G(\frac{q_1^Q(x^c)}{Q})}.$$

Step 3 Next, I will show that for all $x \geq x^c$, $W'(x_1) \leq 0$.

$$\begin{aligned} W'(x_1) = &- K'(x_1)\left[\int_{x^c}^1 (1-F(x))dx - \int_{x^c}^{x_1} xf(x)dx\right] \\ &- (Q-K(x_1))(1-F(x_1)) - (Q-K(x_1))[x_1 f(x_1) - (1-F(x_1))] \end{aligned}$$

Because for all $x_1 \geq x^c$, we have $x_1 f(x_1) - (1-F(x_1)) \geq 0$, so

$$W'(x_1) \leq -K'(x_1)\left[\int_{x^c}^1 (1-F(x))dx - \int_{x^c}^{x_1} xf(x)dx\right] - (Q-K(x_1))(1-F(x_1)).$$

Therefore, a sufficient condition for $W'(x_1) \leq 0$ for all $x_1 \geq x^c$ is that

$$-K'(x_1)\left[\int_{x^c}^1 (1-F(x))dx - \int_{x^c}^{x_1} xf(x)dx\right] \leq (Q-K(x_1))(1-F(x_1))$$

which is equivalent to

$$\frac{x_1(1-F(x_1))}{\int_{x_1}^1 (1-F(x))dx - \int_{x^c}^{x_1} xf(x)dx} \geq \frac{1-G(\frac{q_1^Q(x_1)}{Q})}{1-(1-x_1)G(\frac{q_1^Q(x_1)}{Q})}.$$

Because

$$\frac{d\left[x_1(1-F(x_1)) - \int_{x_1}^{1}(1-F(x))dx + \int_{x^c}^{x_1} xf(x)dx\right]}{dx_1} = 2(1-F(x_1)) \geq 0,$$

we have

$$x_1(1-F(x_1)) - \int_{x_1}^{1}(1-F(x))dx + \int_{x^c}^{x_1} xf(x)dx \geq x^c(1-F(x^c)) - \int_{x^c}^{1}(1-F(x))dx \geq 0.$$

Further I get for all $x_1 \geq x^c$,

$$\frac{x_1(1-F(x_1))}{\int_{x_1}^{1}(1-F(x))dx - \int_{x^c}^{x_1} xf(x)dx} \geq 1 \geq \frac{1-G(\frac{q_1^Q(x_1)}{Q})}{1-(1-x_1)G(\frac{q_1^Q(x_1)}{Q})},$$

which means for all $x \geq x^c$, $W'(x) \leq 0$. So all $x_1 \in (x^c, 1]$ is dominated by $x_1 = x^c = x_0$. Contradiction.

Step 4 We cannot have $x_1 > x_0$ when (12) does not hold. Thus (12) is a necessary condition for $x_1 > x_0$.

F Examples

If both F and G are uniformly distributed between 0 and 1, we have $x^c = \frac{1}{2}$, thus

$$\frac{x^c(1-F(x^c))}{\int_{x^c}^{1}(1-F(x))dx} = \frac{1-F(x^c)}{\frac{\int_{x^c}^{1}(1-F(x))dx}{1-x^c}} \geq 1.$$

Equation (12) is not satisfied, which further gives us $x_0 = x_1$.

Next I will give an example that satisfies the equation (11). G is still uniformly distributed from 0 to 1, and F is a truncated lognormal distribution as follows.

$$\log(10Y) \sim N(-1, 1)$$

$$F(x) = \begin{cases} Prob(Y \leq x), & x < 1 \\ 1, & x = 1 \end{cases} \tag{29}$$

$$x^c = 0.0490$$

$$\frac{x^c(1-F(x^c))}{\int_{x^c}^1 (1-F(x))dx} = 0.7054$$

We can solve for $q_1^Q(x)$ as follows

$$q_1^Q(x) = \frac{Q}{1+\sqrt{x}},$$

and then we get

$$\frac{1-G(\frac{q_1^Q(x^c)}{Q})}{1-(1-x^c)G(\frac{q_1^Q(x^c)}{Q})} = 0.8188$$

We can also use other commonly used heavy-tailed distributions, such as the log-logistic distribution.

$$\log(20Y) \sim Logistic(0, 0.5)$$
$$F(x) = \begin{cases} Prob(Y \le x), & x < 1 \\ 1, & x = 1 \end{cases} \tag{30}$$

And we can get

$$\frac{x^c(1-F(x^c))}{\int_{x^c}^1 (1-F(x))dx} = 0.6796$$
$$\frac{1-G(\frac{q_1^Q(x^c)}{Q})}{1-(1-x^c)G(\frac{q_1^Q(x^c)}{Q})} = 0.8173$$

I also provide an example here that can be calculated by hand. G is still uniformly distributed from 0 to 1, and F is as follows.

$$F(x) = \begin{cases} 1-(1-x)^n, & x \le \frac{1}{n+1} \\ 1-\frac{A}{x}+\varepsilon(x-\frac{1}{n+1}), & x \in (\frac{1}{n+1}, 1) \\ 1, & x = 1 \end{cases}$$

where $A = \frac{n^n}{(n+1)^{n+1}}$. Thus we can get

$$f(x) = \begin{cases} n(1-x)^{n-1}, & x \leq \dfrac{1}{n+1} \\ \dfrac{A}{x^2} + \varepsilon, & x \in (\dfrac{1}{n+1}, 1) \\ +\infty, & x = 1 \end{cases}$$

and

$$\frac{xf(x)}{1-F(x)} = \begin{cases} \dfrac{nx}{1-x}, & x \leq \dfrac{1}{n+1} \\ \dfrac{\frac{A}{x} + \varepsilon x}{\frac{A}{x} - \varepsilon(x - \frac{1}{n+1})}, & x \in (\dfrac{1}{n+1}, 1) \\ +\infty, & x = 1 \end{cases}$$

Easy to check that $\frac{xf(x)}{1-F(x)}$ increases in x, so that it satisfies assumption 1. We then get $x^c = \frac{1}{n+1}$, and the right-hand-side side of (11) is

$$\begin{aligned} \frac{x^c(1-F(x^c))}{\int_{x^c}^1 (1-F(x))dx} &= \frac{A}{\int_{x^c}^1 \left[\frac{A}{x} - \varepsilon\left(x - \frac{1}{n+1}\right)\right] dx} \\ &= \frac{A}{A\ln(x) - \frac{\varepsilon}{2}x^2 + \frac{\varepsilon}{n+1}x\Big|_{x^c}^1} \\ &= \frac{1}{\ln(n+1) - \frac{\varepsilon}{2A}(\frac{1}{n+1} - 1)^2}. \end{aligned}$$

We can pick $\varepsilon = \frac{A}{n^2}$ and $n = 4$, then this term is approximately equal to 0.62915.

Next we count the left-hand-side term. First we solve for $q_1^Q(x)$ as follows

$$q_1^Q(x) = \frac{Q}{1+\sqrt{x}}.$$

Thus

$$\frac{1 - G(\frac{q_1^Q(x^c)}{Q})}{1 - (1-x^c)G(\frac{q_1^Q(x^c)}{Q})} = \frac{1 - \frac{1}{1+\sqrt{x^c}}}{1 - \frac{1-x^c}{1+\sqrt{x^c}}} = 0.69098 > 0.62915.$$

So we have $x_1 > x_0$.